



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

ANA PAULA MIRANDA DINIZ

**MODELOS DE PREVISÃO DO CONTEÚDO DE SILÍCIO NO FERRO-  
GUSA USANDO REDES NEURAIS ARTIFICIAIS**

VITÓRIA

2018

ANA PAULA MIRANDA DINIZ

## **MODELOS DE PREVISÃO DO CONTEÚDO DE SILÍCIO NO FERRO- GUSA USANDO REDES NEURAIS ARTIFICIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica – PPGEE, do Centro Tecnológico da Universidade Federal do Espírito Santo - UFES, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. José Leandro Felix Salles

VITÓRIA

2018

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Setorial Tecnológica,  
Universidade Federal do Espírito Santo, ES, Brasil)  
Sandra Mara Borges Campos – CRB-6 ES-000593/O

---

Diniz, Ana Paula Miranda, 1991-  
D585m Modelos de previsão do conteúdo de silício no ferro-gusa  
usando redes neurais artificiais / Ana Paula Miranda Diniz. –  
2018.  
173 f. : il.

Orientador: José Leandro Felix Salles.  
Dissertação (Mestrado em Engenharia Elétrica) –  
Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Altos-fornos. 2. Silício. 3. Previsão. 4. Redes neurais  
(Computação). 5. Wavelets (Matemática). 6. Algoritmo de poda  
(Computação). I. Salles, José Leandro Felix. II. Universidade  
Federal do Espírito Santo. Centro Tecnológico. III. Título.

CDU: 621.3

---

**ANA PAULA MIRANDA DINIZ**

**MODELOS DE PREVISÃO DO CONTEÚDO DE SILÍCIO NO FERRO-  
GUSA USANDO REDES NEURAIAS ARTIFICIAIS**

Dissertação submetida ao programa de Pós-Graduação em Engenharia Elétrica – PPGEE, do Centro Tecnológico da Universidade Federal do Espírito Santo – UFES, como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica.

Aprovada em 23 de Março de 2018.

**BANCA EXAMINADORA**

---

**Prof. Dr. José Leandro Felix Salles – Orientador**  
**Universidade Federal do Espírito Santo - UFES**

---

**Prof. Dr. Patrick Marques Ciarelli – Membro Interno**  
**Universidade Federal do Espírito Santo - UFES**

---

**Prof. Dr. Marcelo Lucas Pereira Machado – Membro Externo**  
**Instituto Federal do Espírito Santo – IFES Vitória**

*“No meio da dificuldade encontra-se a oportunidade”  
(Albert Einstein)*

*À Deus e aos meus pais,  
com todo o meu amor e gratidão.*

## **AGRADECIMENTOS**

Primeiramente, agradeço à Deus por iluminar o meu caminho e me dar forças para seguir sempre em frente. Dedico esta dissertação aos meus pais, Marco Antônio e Maria Elvira, por serem os meus maiores incentivadores, responsáveis pela minha educação base e que nunca mediram esforços para que eu chegasse até esta etapa de minha vida.

Agradeço ao Professor Dr. José Leandro Félix Sallles, pelos ensinamentos, orientação e, principalmente, pela confiança. Ao Professor Dr. Klaus Fabian Coco, que sempre me auxiliou nas incontáveis dúvidas que surgiram ao longo deste trabalho.

Pelo suporte e concessão da bolsa de estudos, agradeço à CAPES e ao PPGEE. Por fim, agradeço à ArcelorMittal Tubarão, especialmente ao Engenheiro Douglas Ruy e toda a sua equipe, pelo incentivo e oportunidade no desenvolvimento deste projeto.

## RESUMO

O teor de silício presente no ferro-gusa tem sido usado ao longo dos anos como um dos índices mais representativos do estado térmico de um alto-forno, bem como da qualidade do ferro-gusa que está sendo produzido. Trabalhos anteriores comprovam a eficiência das Redes Neurais Artificiais (RNA) no que concerne a previsão do teor de silício. Partindo desta premissa, este trabalho propõe o uso de modelos neurais para prever os elementos da série temporal do teor de silício no ferro-gusa, que devido aos atrasos de atualização das medições do teor de silício, serão utilizados horizontes de previsão de no mínimo 3 horas à frente e de no máximo 8 horas à frente. Serão testados modelos com e sem entradas exógenas, utilizando para a seleção das mesmas, a expertise dos operadores de alto-forno e técnicas de otimização como o Algoritmo de Poda. Além disto, visando melhorar o desempenho das previsões, será feita a decomposição do sinal da série de silício em componentes aditivas usando a *Maximal Overlap Discrete Wavelet Packet Transform* (MODWPT) e, posteriormente, tais subséries serão aplicadas como entradas de modelos neurais. Os resultados obtidos com dados reais de uma indústria siderúrgica indicam que o algoritmo híbrido apresenta resultados superiores aos dos modelos usando RNA somente com entradas exógenas. Portanto, os resultados obtidos por este trabalho poderão antecipar ações de controle durante o processo produtivo, contribuindo não só na qualidade do produto final como também para a redução dos custos associados à sua produção.

**Palavras-chave:** Alto-Forno, Silício, Previsão, Redes Neurais Artificiais, Algoritmo de Poda, Transformada *Wavelet*.



## ABSTRACT

The silicon content present in hot metal has been used over the years as one of the most representative indices of the thermal state of a blast furnace, as well as the quality of pig iron produced. Previous work has demonstrated the efficiency of Artificial Neural Networks (ANN) in terms of prediction of silicon content. Based on this premise, this work proposes the use of neural models to predict the elements of the time series of the silicon content in the hot metal, which due to delays in updating the silicon content measurements will be used forecast horizons of at least 3 hours ahead and a maximum of 8 hours ahead. Models with and without exogenous entrances will be tested, using for their selection, the expertise of the blast furnace operators and optimization techniques such as the Pruning Algorithm. In addition, in order to improve prediction performance, the signal of the silicon series will be decomposed into additive components using the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT), and later these subseries will be applied as inputs to neural models. The results indicate that the hybrid algorithm presents superior performance to the algorithm using only ANN. The results obtained with real data of a steel industry indicate that the hybrid algorithm presents superior results to the models using ANN only with exogenous entrances. Therefore, the results obtained by this work can anticipate control actions during the production process, contributing not only to the quality of the final product but also for the reduction of the costs associated with its production.

**Keywords:** Blast Furnace, Silicon, Forecasting, Artificial Neural Networks, Pruning Algorithm, Wavelet Transform.

## LISTA DE FIGURAS

Figura 1 – Exemplos de séries temporais não-estacionárias com 200 amostras no tempo: a) Processo estacionário e <i>trend</i> -estacionário; b) Processo com mudança de nível; c) Processo com mudança de variância; d) Processo com raízes unitárias (caminhada aleatória) .....	32
Figura 2 – O uso de janelas deslizantes na análise local do tempo .....	35
Figura 3 – Diagrama do plano tempo-frequência pela análise da STFT .....	36
Figura 4 – Diagrama do plano tempo-frequência pela análise da WT.....	37
Figura 5 – Exemplo de uma <i>wavelet</i> de Haar.....	38
Figura 6 – Representação das funções da família Daubechies de db2 até db10 .....	39
Figura 7 – Representação das funções da família Symlets de sym2 até sym8. ....	40
Figura 8 – Representação das funções da família Coiflets de coif1 até coif5 .....	40
Figura 9 – Processo de decomposição do sinal via DWT. ....	42
Figura 10 – Processo de decomposição do sinal via DWPT .....	44
Figura 11 – Processo de decomposição do sinal via MODWPT .....	47
Figura 12 – Sobreposição das bandas de frequência de um processo MODWT .....	48
Figura 13 – Componentes do neurônio biológico .....	50
Figura 14 – Modelo MCP .....	51
Figura 15 – Redes Neurais Artificiais do tipo <i>Feedforward</i> : a) Rede de Camada Única b) Rede com Múltiplas Camadas. Os níveis de bias foram omitidos por conveniência de representação. ....	52
Figura 16 – Exemplo de Rede Recorrente com neurônios ocultos. Os níveis de bias foram omitidos por conveniência de representação .....	53
Figura 17 – Diagrama de blocos da aprendizagem supervisionada.....	55
Figura 18 – Fluxo de processamento do Algoritmo de <i>Back-Propagation</i> .....	56
Figura 19 – Exemplo de atrasadores formando uma janela de tempo de comprimento $n_w$ na entrada de uma RNA. ....	58
Figura 20 – Arquitetura genérica de uma TDNN com uma única camada oculta .....	59
Figura 21 – Rede NARX com uma única camada oculta, capaz de prever uma saída um passo à frente, configurada no modo de operação paralelo .....	60
Figura 22 – Rede NARX com uma única camada oculta, capaz de prever uma saída um passo à frente, configurada no modo de operação série-paralelo.....	61
Figura 23 – Exemplo de uma rede NAR com 6 neurônios na camada oculta e 4 neurônios de entrada para prever uma determinada saída 3 passos à frente.....	62

Figura 24 – Critério de parada de validação cruzada: simulação.....	64
Figura 25 – Alto-Forno 3 da ArcelorMittal Tubarão: a) Vista aérea; b) Torre do alto-forno.....	69
Figura 26 – Exemplo de amostras ausentes devido as falhas de medição: a) todas as variáveis possuem amostras ausentes; b) nem todos os parâmetros possuem dados faltantes .....	91
Figura 27 – Exemplo de identificação de um período de “não-normalidade” .....	92
Figura 28 – Identificação de períodos contendo <i>outliers</i> .....	93
Figura 29 – Identificação de 4 dos 10 subconjuntos de dados.....	94
Figura 30 – Eliminação das 6 amostras no início e no fim de cada subconjunto, representado por quatro séries temporais com 200 amostras cada: a) representação dos subconjuntos; b) evidenciadas as 6 amostras, no início e no fim de cada subconjunto, a serem eliminadas de cada variável; c) ilustrados os novos pontos de início e fim dos subconjuntos, após a eliminação das amostras. ....	95
Figura 31 – Inserção por interpolação de 6 amostras entre os subconjuntos: a) apresentação das novas amostras inseridas em cada série através de uma interpolação; b) ilustração das respectivas séries resultantes .....	96
Figura 32 – Comparação entre médias e variâncias de cinco pacotes de dados da série de teor de silício .....	98
Figura 33 – Comportamento da série de Teor de Silício em relação à linearidade .....	99
Figura 34 – Estrutura da topologia de rede NIO aplicada neste trabalho .....	100
Figura 35 – Estrutura da topologia de rede NAR aplicada neste trabalho.....	100
Figura 36 – Estrutura da topologia de rede NARX aplicada neste trabalho .....	101
Figura 37 – O MSE médio de treinamento obtido para cada tamanho de camada oculta.....	101
Figura 38 – Previsão do Teor de Silício do ferro-gusa: a) Rede NIO com horizonte de previsão de 1 hora; b) Rede NARX com horizonte de previsão de 1 hora; c) Rede NAR com horizonte de previsão de 1 hora; d) Rede NIO com horizonte de previsão de 3 horas; e) Rede NARX com horizonte de previsão de 3 horas; f) Rede NAR com horizonte de previsão de 3 horas. ....	102
Figura 39 – Critérios de Desempenho versus Horizonte de Previsão: a) MSE; b) MAPE (%); c) PEA (%) .....	104
Figura 40 – Previsão do Teor de Silício do ferro-gusa empregando uma Rede NIO e Filtro a) Média Móvel e horizonte de previsão de 1 hora; b) Mediana Móvel e horizonte de previsão de 1 hora; c) <i>Wavelet</i> e horizonte de previsão de 1 hora; d) Média Móvel e horizonte de previsão de 3 horas; e) Mediana Móvel e horizonte de previsão de 3 horas; f) <i>Wavelet</i> e horizonte de previsão de 3 horas .....	108
Figura 41 – Previsão do Teor de Silício do ferro-gusa empregando uma Rede NARX e Filtro a) Média Móvel e horizonte de previsão de 1 hora; b) Mediana Móvel e horizonte de previsão de 1 hora; c) <i>Wavelet</i> e horizonte de previsão	

de 1 hora; d) Média Móvel e horizonte de previsão de 3 horas; e) Mediana Móvel e horizonte de previsão de 3 horas; f) <i>Wavelet</i> e horizonte de previsão de 3 horas .....	109
Figura 42 – Previsão do Teor de Silício do ferro-gusa empregando uma Rede NAR e Filtro a) Média Móvel e horizonte de previsão de 1 hora; b) Mediana Móvel e horizonte de previsão de 1 hora; c) <i>Wavelet</i> e horizonte de previsão de 1 hora; d) Média Móvel e horizonte de previsão de 3 horas; e) Mediana Móvel e horizonte de previsão de 3 horas; f) <i>Wavelet</i> e horizonte de previsão de 3 horas .....	110
Figura 43 – Rede NIO resultante após aplicação do Algoritmo de Poda .....	114
Figura 44 – Rede NARX resultante após aplicação do Algoritmo de Poda.....	115
Figura 45 – Previsão do Teor de Silício do ferro-gusa aplicando Algoritmo de Poda: Rede NIO com horizonte de previsão de a) 1 hora; b) 3 horas; c) 6 horas; Rede NARX com horizonte de previsão de d) 1 hora; e) 3 horas; f) 6 horas.....	116
Figura 46 – Previsão do Teor de Silício do ferro-gusa aplicando Algoritmo de Poda: Rede NAR com horizonte de previsão de a) 1 hora; b) 3 horas; c) 6 horas.....	119
Figura 47 – A MODWPT: Série decomposta em 8 subséries .....	121
Figura 48 – Estrutura do algoritmo MODWPT-NAR .....	122
Figura 49 – O MSE médio de treinamento obtido para cada tamanho de camada oculta do modelo híbrido MODWPT-NAR.....	123
Figura 50 – Previsão do Teor de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NAR com horizonte de previsão de a) 1 hora; b) 2 horas; c) 3 horas; d) 4 horas.....	125
Figura 51 – Previsão do Teor de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NAR com horizonte de previsão de a) 5 horas; b) 6 horas; c) 7 horas; d) 8 horas.....	126
Figura 52 – MODWPT-NIO: a) Estrutura do algoritmo; b) Implementação da RNA do tipo NIO .....	128
Figura 53 – O MSE médio de treinamento obtido para cada tamanho de camada oculta do modelo híbrido MODWPT-NIO .....	129
Figura 54 – Previsão do Teor de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NIO com horizonte de previsão de a) 1 hora; b) 2 horas; c) 3 horas; d) 4 horas.....	130
Figura 55 – Previsão do Teor de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NIO com horizonte de previsão de a) 5 horas; b) 6 horas; c) 7 horas; d) 8 horas.....	131
Figura 56 – Representação do Alto-Forno com suas entradas e saídas.....	145
Figura 57 – Zonas internas do Alto-Forno .....	146
Figura 58 – Vazamento de gusa e escória do cadinho pelo furo de gusa e a separação ao passar pela escumadeira em uma casa de corrida .....	150

Figura 59 – Padrão típico de operação de vazamento de um alto-forno com 4 furos de gusa, em que os furos 1 e 3 são operados de forma alternativa e sem atrasos .....	151
Figura 60 – Representação dos equipamentos e indicadores fundamentais para o controle do processo.....	153
Figura 61 – Filtro linear, com entrada $a_t$ , saída $Z_t$ e função de transferência $\psi(B)$ .....	160

## LISTA DE TABELAS

Tabela 1 – Literaturas de Referência: Resumo .....	77
Tabela 2 – Entradas empregadas nos modelos implementados nas literaturas de referência .....	80
Tabela 3 – Potenciais variáveis de entrada do modelo e suas principais características .....	82
Tabela 4 – Resultado do Teste de Estacionariedade .....	97
Tabela 5 – Resultados dos Testes de Linearidade .....	98
Tabela 6 – Critérios de desempenho obtidos pelas previsões .....	103
Tabela 7 – Comparativo dos critérios de desempenho das previsões aplicando Filtros de Média Móvel, Mediana Móvel e <i>Wavelet</i> aos dados da Rede NIO .....	105
Tabela 8 – Comparativo dos critérios de desempenho das previsões aplicando Filtros de Média Móvel, Mediana Móvel e <i>Wavelet</i> aos dados da Rede NARX.....	106
Tabela 9 – Comparativo dos critérios de desempenho das previsões aplicando Filtros de Média Móvel, Mediana Móvel e <i>Wavelet</i> aos dados da Rede NAR .....	106
Tabela 10 – Ocorrência das entradas e ordem dos regressores resultantes após cada implementação do Algoritmo de Poda aplicado às Rede NIO e NARX.....	112
Tabela 11 – Comparativo dos critérios de desempenho das previsões antes e depois da aplicação do Algoritmo de Poda nas Redes NIO e NARX.....	117
Tabela 12 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NAR .....	118
Tabela 13 – Comparativo dos critérios de desempenho das previsões antes e depois da aplicação do Algoritmo de Poda na Rede NAR.....	119
Tabela 14 – Tamanho Da Camada Oculta da rede para a Previsão de cada uma das Subséries .....	124
Tabela 15 – Comparativo entre Critérios de Desempenhos das Previsões de 1 a 8 passos à frente do Modelo MODWPT-NAR e do Modelo NAR.....	127
Tabela 16 – Comparativo entre Critérios de Desempenhos das Previsões de 1 a 8 passos à frente dos Modelos MODWPT-NAR e MODWPT-NIO .....	132
Tabela 17 – Comparativo entre os resultados gerados pelas diferentes metodologias encontradas na literatura .....	133
Tabela 18 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NIO .....	163
Tabela 19 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NARX.....	168

## LISTA DE ABREVIATURAS

ADF	<i>Augmented Dickey-Fuller</i>
AG	Algoritmo Genético
ARIMA	Autorregressivos Integrados e de Médias Móveis
ARIMAX	Autorregressivos Integrados e de Médias Móveis com Entradas Exógenas
ARMA	Autorregressivos e de Médias Móveis
ARMAX	Autorregressivos de Médias Móveis com Entradas Exógenas
BFG	<i>Blast Furnace Gas</i>
DDS	<i>Dynamic Data System</i>
DF	<i>Dickey-Fuller</i>
DSP	<i>Digital Signal Processing</i>
DWPT	<i>Discrete Wavelet Packet Transform</i>
DWT	<i>Discrete Wavelet Transform</i>
EMD	<i>Empirical Mode Decomposition</i>
EMF	<i>ElectroMotive Force</i>
FIR	<i>Finite Impulse Response</i>
FPE	<i>Final Prediction Error</i>
FT	<i>Fourier Transform</i>
HPF	<i>High-Pass Filter</i>
IFGCN	Índice de Fluxo Gasoso Central Norte
IFGCS	Índice de Fluxo Gasoso Central Sul
IFGPN	Índice de Fluxo Gasoso Periférico Norte
IFGPS	Índice de Fluxo Gasoso Periférico Sul
KPSS	Kwiatkowski, Phillips, Schmidt, and Shin
LPH	<i>Low-Pass Filter</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MLP	<i>Multilayer Perceptron</i>
MLR	<i>Multiple Linear Regression</i>
MODWPT	<i>Maximal Overlap Discrete Wavelet Packet Transform</i>
MODWT	<i>Maximal Overlap Discrete Wavelet Transform</i>
MRA	<i>MultiResolution Analysis</i>

MSE	<i>Mean Squared Error</i>
NAR	<i>Nonlinear AutoRegressive model</i>
NARX	<i>Nonlinear AutoRegressive model with eXogenous inputs</i>
NIO	<i>Nonlinear Input-Output model</i>
OBD	<i>Optimal Brain Damage</i>
OBS	<i>Optimal Brain Surgeon</i>
PEA	Percentual de Erro Absoluto
PLIM	Programação Linear Inteira Mista
PP	Phillips-Perron
QF	<i>Quadrature Filter</i>
RNA	Rede Neural Artificial
STFT	<i>Short-Time Fourier Transform</i>
TDNN	<i>Time Delay Neural Networks</i>
VPG	Velocidade de Produção de Gusa
WT	<i>Wavelet Tranform</i>



## LISTA DE SÍMBOLOS

$Z_t$	Série Temporal no instante $t$
$N$	Número de amostras de uma Série Temporal
$\Delta$	Operação de diferença
$\Delta Z_t$	Primeira diferença de uma Série Temporal
$\Delta^2 Z_t$	Segunda diferença de uma Série Temporal
$\Delta^d Z_t$	$d$ -ésima diferença de uma Série Temporal
$d$	Ordem da diferença
$t$	Instante de Tempo
$H_0$	Hipótese Nula
$H_1$	Hipótese Alternativa
$N(\mu, \sigma^2)$	Distribuição Normal com Média $\mu$ e Variância $\sigma^2$
$\psi(B)$	Função de Transferência
$a_t$	Ruído Branco
$B$	Operador de atraso
$q$	Ordem do termo de Média Móvel
$p$	Ordem do termo Autorregressivo
$\theta_q(B)$	Polinômio de grau $q$ que representa o operador de Média Móvel
$\phi_p(B)$	Polinômio de grau $p$ que representa o operador Autorregressivo
$MA(q)$	Modelo de Média Móvel
$AR(p)$	Modelo Autorregressivo
$ARMA(p, q)$	Modelo ARMA
$\phi_p$	Parâmetro autorregressivo relativo à amostra $Z_{t-p}$
$\theta_q$	Parâmetro de média móvel referente ao ruído $a_{t-q}$ .
$W_t$	Série Temporal diferenciada e estacionária
$ARIMA(p, d, q)$	Modelo ARIMA
$ARMAX(p, q, r)$	Modelo ARMAX
$r_x$	Ordem da equação que representa a variável exógena
$\beta_{r_x}$	Coefficiente da Entrada Exógena
$x(t)$	Sinal no tempo
$\mathcal{F}[x(t)]$	Transformada de Fourier de um sinal $x(t)$

$f_s$	Frequência do sinal $x(t)$
$STFT(\tau, f)$	Transformada de Fourier de Tempo-Curto de um sinal $x(t)$
$g(t - \tau)$	Função responsável pelo janelamento centrado em $\tau$
$\Delta f_s$	Resolução em Frequência
$\Delta t$	Resolução no Tempo
$CWT(a, b)$	Transformada <i>Wavelet</i> Contínua
$a$	Parâmetro de escala
$b$	Parâmetro de translação
$\psi_{a,b}(t)$	Função <i>wavelet</i> base $\psi(t)$ , escalada e deslocada
$N_o$	Ordem da família de <i>wavelets</i>
$dbN_o$	<i>Wavelets</i> da família Daubechies de ordem $N_o$
$symN_o$	<i>Wavelets</i> da família Symlets de ordem $N_o$
$coifN_o$	<i>Wavelets</i> da família Coiflets de ordem $N_o$
$\psi_{j,r}(t)$	Função da família de <i>wavelets</i>
$j$	Parâmetro de dilatação (nível)
$r$	Parâmetro de translação
$a_0$	Variação da dilatação
$b_0$	Fator de Translação
$cD$	Coeficiente de Detalhe
$cA$	Coeficiente de Aproximação
$\phi_{j,r}(t)$	Função de Escalonamento
$J$	Escala Máxima
$a_{j,r}$	Componentes de aproximação na escala máxima
$d_{j,r}$	Coeficientes de detalhe em todas as escalas
$L$	Nível de decomposição máximo
$c\tilde{A}_{j,m}$	$m$ -ésimo elemento do $j$ -ésimo nível dos coeficientes de aproximação
$c\tilde{D}_{j,m}$	$m$ -ésimo elemento do $j$ -ésimo nível dos coeficientes de detalhe
$\tilde{g}_{j,l}$	Coeficientes do filtro passa-baixa obtidos a partir do filtro $g_l$
$\tilde{h}_{j,l}$	Coeficientes do filtro passa-alta obtidos a partir do filtro $h_l$
$g_l$	Coeficientes dos filtros escala (passa-baixa)
$h_l$	Coeficientes dos filtros <i>wavelet</i> (passa-alta)

$L_{cf}$	Número de coeficientes dos filtros escala e <i>wavelet</i>
$n$	Número de entradas
$\varphi(\cdot)$	Função de ativação do neurônio
$v$	Número do neurônio
$u_v$	Saída do combinador linear do neurônio $v$
$x_j$	$j$ -ésimo sinal de entrada do neurônio
$w_{vj}$	Peso sináptico do $j$ -ésimo sinal de entrada do $v$ -ésimo neurônio
$b_v$	Limiar correspondente ao neurônio $v$ ( <i>bias</i> )
$y_v$	Sinal de saída do neurônio $v$
$\mathbf{w}(t)$	Vetor de Pesos sinápticos no instante $t$
$\mathbf{x}(t)$	Vetor de sinais de entrada
$\Delta \mathbf{w}(t)$	Ajuste de peso
$y(t)$ ou $y_t$	Resposta Real
$\hat{y}(t)$ ou $\hat{y}_t$	Resposta Prevista
$e(t)$	Sinal de Erro
$\varepsilon(t)$	Função Custo
$n_T$	Número de exemplos de treinamento
$e^2(t)$	Erro quadrático da saída
$\mathbf{x}_{in}(t)$	Padrão de entrada
$n_w$	Comprimento da janela temporal
$k$	Horizonte de Previsão
$n_a$	Número de atrasos da memória de saída
$n_b$	Número de atrasos da memória de entrada
$y_d(t + k)$	Resposta desejada $k$ passos à frente
$\mathbf{g}(\mathbf{w})$	Vetor de gradiente calculado para o ponto de operação $\mathbf{w}$
$\mathbf{H}(\mathbf{w})$	Matriz Hessiana calculado para o ponto de operação $\mathbf{w}$
$\Delta \varepsilon$	Variação do erro
$\lambda$	Multiplicador de Lagrange
$S$	Lagrangiano
$x_i$	Valor pontual de uma determinada variável
$x_{min}$	Valor Mínimo assumido por uma variável
$x_{max}$	Valor Máximo assumido por uma variável
$x_{norm}$	Valor normalizado de uma variável

$X_{WP}$

Subsérie do sinal  $x(t)$  gerada através da Transformada *Wavelet*

$\hat{X}_{WP}$

Subsérie Prevista

# SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>23</b>
1.1 OBJETIVOS DO TRABALHO.....	25
1.2 ORGANIZAÇÃO DA DISSERTAÇÃO .....	26
<b>2 ANÁLISE DE SINAIS.....</b>	<b>28</b>
2.1 SÉRIES TEMPORAIS .....	29
<b>2.1.1 Séries temporais não-estacionárias .....</b>	<b>30</b>
2.1.1.1 Principais tipos de não-estacionariedade .....	31
<b>2.1.2 Testes de Estacionariedade .....</b>	<b>32</b>
<b>2.1.3 Testes de Linearidade.....</b>	<b>33</b>
2.2 A TRANSFORMADA <i>WAVELET</i> .....	34
<b>2.2.1 A Base <i>Wavelet</i> de Haar .....</b>	<b>38</b>
<b>2.2.2 Bases <i>Wavelets</i> de Daubechies .....</b>	<b>38</b>
<b>2.2.3 Bases <i>Wavelets</i> de Symlets e Coiflets .....</b>	<b>39</b>
<b>2.2.4 Transformada <i>Wavelet</i> Discreta.....</b>	<b>40</b>
2.2.4.1 A <i>Wavelet</i> Packet Discreta .....	43
2.3 TRANSFORMADA <i>WAVELET</i> APLICADA NA ANÁLISE DE SÉRIES TEMPORAIS.....	45
<b>3 REDES NEURAIS ARTIFICIAIS.....</b>	<b>49</b>
3.1 O MODELO DE NEURÔNIO .....	49
3.2 PRINCIPAIS TIPOS DE ARQUITETURAS.....	52
3.3 ALGORITMOS DE APRENDIZAGEM .....	54
3.4 REDES NEURAIS COM ATRASO DE TEMPO .....	57
3.5 O PROJETO DE UMA RNA.....	62
<b>3.5.1 Medidas de Desempenho .....</b>	<b>64</b>
3.6 ALGORITMOS DE PODA .....	65
<b>3.6.1 Poda da Rede baseada na Matriz Hessiana .....</b>	<b>66</b>
<b>4 ESTUDO DE CASO .....</b>	<b>69</b>
4.1 REVISÃO BIBLIOGRÁFICA.....	70
4.2 PRINCIPAIS VARIÁVEIS ENVOLVIDAS NO PROCESSO .....	81
<b>5 DESENVOLVIMENTO DE MODELOS E RESULTADOS .....</b>	<b>87</b>
5.1 METODOLOGIA .....	87
5.2 REAMOSTRAGEM DOS DADOS .....	90

5.3 TRATAMENTO DOS DADOS .....	90
5.4 NORMALIZAÇÃO DOS DADOS .....	96
5.5 ANÁLISE ESTATÍSTICA DA SÉRIE DE TEOR DE SILÍCIO .....	97
5.6 MODELAGEM E PREVISÃO DO TEOR DE SILÍCIO UTILIZANDO RNA .....	99
5.7 MODELAGEM E PREVISÃO DO TEOR DE SILÍCIO UTILIZANDO UM ALGORITMO HÍBRIDO DE MODWPT E NAR .....	120
5.8 MODELAGEM E PREVISÃO DO TEOR DE SILÍCIO UTILIZANDO UM ALGORITMO HÍBRIDO DE MODWPT E NIO .....	128
5.9 COMPARATIVO ENTRE AS METODOLOGIAS APLICADAS NA LITERATURA .....	133
<b>6 CONCLUSÃO .....</b>	<b>135</b>
<b>REFERÊNCIAS.....</b>	<b>137</b>
<b>APÊNDICE A: PRODUÇÃO DE FERRO-GUSA EM ALTO-FORNO .....</b>	<b>145</b>
A.1 ASPECTOS INTERNOS DO ALTO-FORNO .....	146
A.2 O FUNCIONAMENTO DO ALTO-FORNO .....	148
<b>A.2.1 O vazamento do gusa.....</b>	<b>149</b>
<b>A.2.2 Controle operacional do Alto-forno .....</b>	<b>151</b>
A.3 O TEOR DE SILÍCIO NO FERRO-GUSA .....	155
<b>A.3.1 Mecanismos de transferência do silício .....</b>	<b>156</b>
<b>A.3.2 Efeitos do silício no processo produtivo .....</b>	<b>158</b>
<b>ANEXO A: MODELOS ARIMA E ARIMAX.....</b>	<b>160</b>
<b>ANEXO B: RESULTADOS OBTIDOS APÓS CADA IMPLEMENTAÇÃO DO ALGORITMO DE PODA APLICADO À REDE NIO.....</b>	<b>163</b>
<b>ANEXO C: RESULTADOS OBTIDOS APÓS CADA IMPLEMENTAÇÃO DO ALGORITMO DE PODA APLICADO À REDE NARX.....</b>	<b>168</b>

## 1 INTRODUÇÃO

Visando uma maior competitividade no mercado mundial, as indústrias buscam incessantemente a redução de seus gastos operacionais, aliando uma gestão eficiente às técnicas que permitam o acesso às informações relevantes durante o processo, facilitando e potencializando a tomada de decisões estratégicas (STEARNS, 2007; AGUIRRE, 2015).

Ao longo da história, as usinas siderúrgicas assumiram um papel de grande relevância na economia mundial, impulsionando o processo de industrialização. Geerdes, Toxopeus e Van Der Vliet (2009) afirmam que cerca de dois terços da produção de aço no mundo provêm de usinas siderúrgicas integradas. Estas usinas produzem o ferro-gusa no alto-forno e posteriormente o encaminham para a refino na aciaria, resultando no aço. A qualidade do ferro-gusa produzido é importante para determinar o quão dispendioso será produzir o aço, bem como restringir em que tipos finais de aço o ferro-gusa poderá ser empregado (RIZZO, 2009).

O ferro-gusa, ou gusa, se forma a partir da redução do minério de ferro. Além disto, também são formados, como subprodutos, o gás de alto-forno (*Blast Furnace Gas – BFG*) e a escória, proveniente da aglomeração dos minerais e das cinzas resultantes da queima dos combustíveis (GEERDES; TOXOPEUS; VAN DER VLIET, 2009; MOURAO *et al.*, 2011). Conforme apresentado por Arrivabene, Machado e Sobrinho (2003), levam-se cerca de seis a oito horas para que as matérias-primas carregadas sejam processadas e se acumulem na base do alto-forno, na forma de gusa e escória, separados por diferença de densidade, ficando o gusa na parte inferior.

Os processos na indústria siderúrgica são muitas vezes complexos e não-lineares. Além disto, devido às condições hostis no interior do alto-forno (alta pressão, temperatura e ambiente erosivo) e à constante ocorrência de distúrbios durante o processo, a avaliação do estado interno do processo por medição direta não é uma tarefa trivial (MOURAO *et al.*, 2011; SAXEN; GAO, C.; GAO, Z., 2013). Em vez disso, o estado interno pode ser avaliado baseando-se no comportamento de outras variáveis do processo que podem ser medidas diretamente e que potencialmente afetam no estado do processo (BAG, 2007; CHEN, 2001; DHOND; GUPTA; VADHAVKAR, 2000; SAXEN; GAO, C.; GAO, Z., 2013; TUNÇKAYA; KÖKLÜKAYA, 2016; WANG; LIU, 2011).

Em geral, o controle é realizado pelos operadores de alto-forno, que determinam o estado do processo a partir de variáveis mais fáceis de serem medidas *on-line* por instrumentação padrão. Através destas informações, os operadores realizam ações de controle com o intuito de garantir a normalidade do sistema e controlar a qualidade do produto final, em geral, caracterizada pela temperatura e composição do ferro-gusa. Contudo, este controle é uma tarefa desafiadora devido aos atrasos de tempo, comportamento não-linear, inércia do processo e irregularidade entre os intervalos de abertura e fechamento do furo de gusa (RADHAKRISHNAN; MOHAMED, 2000; SAXEN; GAO, C.; GAO, Z., 2013).

Neste contexto, a identificação de sistemas pode ser uma importante ferramenta para a obtenção de modelos matemáticos capazes de descrever as relações de causa e efeito do sistema em questão (AGUIRRE, 2015). Utilizando dados históricos que descrevem a evolução temporal de uma ou mais variáveis do processo, é possível, então, monitorar e até mesmo prever o estado de funcionamento do alto-forno (SAXEN; GAO, C.; GAO, Z., 2013).

No caso em questão, o teor de silício no ferro-gusa é utilizado por ser um importante indicador do estado térmico de um alto-forno e, conseqüentemente, ser capaz de refletir qual a qualidade do aço e o quão dispendioso será produzi-lo (GEERDES; TOXOPEUS; VAN DER VLIET, 2009; MOURAO *et al.*, 2011). Os valores elevados do teor de silício indicam um aumento do aporte térmico do forno e, em alguns casos, pode refletir em um excedente de coque dentro do forno. Assim, uma vez que o custo do coque é dominante na fabricação do ferro-gusa, há benefícios econômicos bastante evidentes no que concerne ao controle mais rigoroso do teor de silício (SAXEN; GAO, C.; GAO, Z., 2013; SAXEN; PETTERSSON, 2007; WANG; LIU, 2011). Na aciaria, o teor de silício no ferro-gusa é fundamental para o fornecimento da energia ao processo, devido à sua oxidação. No entanto, teores muito elevados de silício podem comprometer a produtividade em virtude da constante necessidade de ajustes do processo (CAMPOS, 1983; GUSTAVSSON, 2004; MOURAO *et al.*, 2011). Desta forma, espera-se uma variação pequena do teor de silício entre as corridas, de maneira a otimizar a operação dos conversores da aciaria.

Como forma de assegurar o bom desempenho da produção do aço, a composição físico-química do ferro-gusa produzido deve apresentar-se adequada e com a menor variação possível. Na prática, a composição química do gusa e da escória é medida



via espectromia por raios-x das amostras dos materiais. Entretanto, por se tratar de uma técnica laboratorial, os resultados desta análise não são mensurados de maneira *on-line* como as demais variáveis de controle do alto-forno medidas diretamente (RADHAKRISHNAN; MOHAMED, 2000). De fato, a falta de medições diretas das condições internas torna o problema de controle do teor de silício no ferro-gusa ainda mais desafiador (SAXEN; GAO, C.; GAO, Z., 2013).

Portanto, somando-se o longo tempo transcorrido desde o carregamento da matéria-prima no topo do alto-forno até sua retirada na forma de ferro e o tempo necessário para que os resultados da análise laboratorial da composição química do ferro-gusa, bem como do percentual de silício presente no mesmo, fique à disposição dos operadores, faz com que atitudes necessárias, visando a garantia de um produto final de boa qualidade e com baixo custo, sejam tomadas tardiamente.

Neste contexto, em virtude do grande interesse demonstrado por uma siderúrgica local em aprimorar a qualidade de seu produto final, pretende-se obter uma modelagem capaz de prever o teor de silício no ferro-gusa produzido, visando auxiliar a tomada de decisão para a manutenção do processo. Desta forma, os resultados advindos desta pesquisa possibilitarão não apenas o aumento na qualidade do produto final como também a redução dos custos associados à sua produção.

## 1.1 OBJETIVOS DO TRABALHO

O objetivo geral deste trabalho é a implantação de um sistema de previsão do teor de silício do ferro-gusa baseado em modelos caixa-preta, isto é, uma modelagem que possa descrever as relações de causa e efeito do processo produtivo, sem a necessidade de profundos conhecimentos a respeito da dinâmica que rege as complexas reações que ocorrem no interior do alto-forno.

Este estudo parte da premissa de que, devido às condições hostis do interior do alto-forno e à constante ocorrência de instabilidades operacionais, a avaliação do estado interno deste processo, por medição direta, além de dispendioso, não é uma tarefa trivial. Portanto, tendo em vista que o teor de silício é um importante indicador do estado térmico, através desta variável e demais fatores que influenciam no processo, será possível obter uma modelagem capaz de refletir não apenas o estado interno do alto-forno, como também a qualidade do aço e os custos associados à sua produção. Os objetivos específicos desta pesquisa são:

- Determinar diferentes estruturas de modelos não-lineares empregado as redes neurais com atrasos de tempo;
- Analisar a estacionariedade e linearidade da série temporal do teor de silício;
- Testar conjuntos de entradas exógenas e os seus respectivos atrasos de tempo determinados através de ferramentas de otimização, auxiliado pelo conhecimento técnico da dinâmica do processo;
- Implementar uma decomposição da série de silício, de maneira que possam ser utilizados sinais mais simples como entradas do modelo, visando um aumento do desempenho da previsão;
- Gerar previsões de até oito horas à frente para o teor do silício no ferro-gusa com desempenho satisfatório.

Deve-se frisar que, devido à grande complexidade intrínseca aos modelos que utilizam entradas exógenas, o principal objetivo deste trabalho é propor uma metodologia que utilize apenas as observações autorregressivas do teor de silício no ferro-gusa para prever os valores futuros. Para tanto, será proposta uma ferramenta híbrida empregando técnicas de decomposição e Redes Neurais Artificiais (RNA).

## 1.2 ORGANIZAÇÃO DA DISSERTAÇÃO

O texto está organizado da seguinte forma: no Capítulo 2 será feita uma breve descrição da análise de séries temporais através da identificação linear paramétrica. Neste capítulo, também será apresentada uma revisão da literatura das decomposições *wavelet*, destacando a utilização da *Maximal Overlap Discrete Wavelet Packet Transform* (MODWPT) no estudo de séries temporais.

No Capítulo 3, as RNA serão apresentadas como ferramenta para a modelagem e previsão do teor de silício no ferro-gusa, descrevendo também a importância da utilização dos Algoritmos de Poda para a seleção de entradas e dos seus respectivos atrasos para o modelo.

No Capítulo 4 será apresentada uma breve descrição dos trabalhos realizados na área, bem como uma descrição do processo de análise do teor de silício no ferro-gusa. No Capítulo 5 será feito um detalhamento das metodologias empregadas para o desenvolvimento do modelo de previsão do teor de silício e realizada a análise e a discussão dos resultados obtidos em cada uma das etapas do trabalho.

No Capítulo 6 serão apresentadas as conclusões deste estudo e também algumas sugestões para o desenvolvimento de trabalhos futuros. Por fim, no Apêndice A será realizada uma descrição sucinta do processo de produção do ferro-gusa no alto-forno, relacionando as variáveis mais importantes e destacando a importância do silício para a análise qualitativa do processo.

## 2 ANÁLISE DE SINAIS

Os sinais são capazes de descrever uma grande variedade de fenômenos físicos. Tratam-se de uma grandeza física variável no tempo que contém algum tipo de informação mensurável que diz respeito ao comportamento de um sistema físico (OPPENHEIM; WILLSKY, 2012). Embora possam ser representados de diversas maneiras, a informação do sinal é intrínseca a algum tipo de variação (HAYKIN; VEEN, 2001).

Em geral, os processos industriais são monitorados através dos mais variados tipos de sensores, que captam os sinais (tais como os de temperatura, pressão e vazão) e os encaminha para sistemas, que recolhem informações e auxiliam o processo de tomada de decisão. Desta forma, o desenvolvimento de técnicas que permitam o monitoramento e a análise destes sinais pode contribuir para a implementação de estratégias de controle em aplicações industriais (DHOND; GUPTA; VADHAVKAR, 2000; KIRCHGÄSSNER; WOLTERS, 2007).

Com o aprimoramento das disciplinas de controle e inteligência computacional, a modelagem de séries temporais passou a ser utilizada para a predição de propriedades a partir de dados da planta (AGUIRRE, 2015). Devido à dependência temporal entre as amostras do mesmo sinal (autocorrelação), a análise de séries temporais permite fazer previsões sobre o futuro a partir da observação do comportamento passado (BARROS, 2014; MORETTIN; TOLOI, 2006).

Neste contexto, o desenvolvimento de ferramentas de previsão que promovam com antecedência as informações referentes ao processo pode permitir que ações de controle sejam tomadas em tempo hábil, visando o aumento na qualidade do produto final, reduzindo as incertezas e garantindo a redução dos custos associados à produção (DHOND; GUPTA; VADHAVKAR, 2000; SAXEN; GAO, C.; GAO, Z, 2013). A análise de sinais constitui uma das ferramentas mais importantes do Processamento Digital de Sinais (*Digital Signal Processing* - DSP), área responsável por interpretar ou manipular um sinal a fim de extrair informações que não estavam disponíveis no formato original ou adequá-lo a uma aplicação específica. Contudo, uma grande variedade dos sinais do mundo real possui características não-estacionárias e, frequentemente, eles estão sujeitos aos ruídos introduzidos em sua captação, que apodem interferir na sua análise (OLIVEIRA, 2007a).

A limitação apresentada pelos métodos na análise de sinais não-estacionários possibilitou o desenvolvimento de técnicas de processamento de sinais tais como a Transformada *Wavelet*. Esta transformada constitui um método de análise capaz de processar dados estacionários e não-estacionários, consolidando-se como uma das ferramentas mais potentes no contexto do DSP. Em virtude da sua capacidade de manipulação do sinal, através das *wavelets* é possível segmentar a série em subséries mais simples, de maneira a facilitar a extração de informações relevantes e a análise do sinal (OLIVEIRA, 2007a).

## 2.1 SÉRIES TEMPORAIS

Uma série temporal é definida como um conjunto de observações ordenadas no tempo, que possuem dependências entre si (MORETTIN; TOLOI, 2006). A análise de séries temporais em busca da detecção de regularidades entre as observações das variáveis que a compõem, de modo a encontrar a lei que as define, torna possível prever observações futuras com certa precisão (KIRCHGÄSSNER; WOLTERS, 2007).

Os modelos utilizados para descrever as séries temporais são processos controlados por leis probabilísticas, isto é, processos estocásticos (MORETTIN; TOLOI, 2006). No início da década de 1920, Evgenij Evgenievich Slutsky e George Udny Yule desenvolveram modelos utilizando médias móveis e processos autorregressivos para representar as séries como somas de processos estocásticos (KIRCHGÄSSNER; WOLTERS, 2007).

Na década de 70, George E. P. Box e Gwilym M. Jenkins, assumiram que havia um modelo estocástico comum para todo o processo de geração de séries temporais, aplicando métodos para implementação destes modelos de maneira empírica. Inicialmente, este método identifica um modelo específico com base em certas configurações estatísticas, a partir de então, os parâmetros deste modelo são estimados e, baseando-se em testes estatísticos, a especificação do modelo é verificada (validação do modelo). Caso o modelo não seja validado, a especificação deve ser alterada e os parâmetros devem ser novamente estimados, de maneira iterativa, até que se encontre um modelo que satisfaça aos critérios pré-estabelecidos e possa ser usado para previsões (BOX; JENKINS; REINSEL, 1976).

Na prática, as ferramentas de previsão de séries temporais podem variar de modelos simples e intuitivos, com pouca análise de dados, a até modelos complexos, sendo necessária uma interpretação mais minuciosa das séries em questão. Para a criação destes modelos, diversos fatores devem ser levados em consideração, tais como o comportamento do fenômeno ou o conhecimento *a priori* da natureza e do objetivo da análise (MORETTIN; TOLOI, 2006).

É importante ressaltar, no entanto, que as previsões obtidas a partir dos modelos tratam-se de uma simplificação da realidade, sendo levados apenas os aspectos essenciais do problema que está sendo tratado (BARROS, 2004) e, portanto, podem ser considerados diferentes tipos de modelos para descrever o comportamento de uma série particular (AGUIRRE, 2015).

É possível classificar os modelos de séries temporais em modelos paramétricos, em que o número de parâmetros é finito, e modelos não-paramétricos, que envolvem um número infinito de parâmetros. Os modelos paramétricos apresentam a sua análise no domínio do tempo. Dentre estes, os mais utilizados são os modelos de erro (ou de regressão), os modelos Autorregressivos e de Médias Móveis (ARMA), os modelos Autorregressivos Integrados e de Médias Móveis (ARIMA), modelos de memória longa, modelos estruturais e modelos não-lineares, utilizando por exemplo, as RNA (MORETTIN; TOLOI, 2006).

Em relação aos modelos não-paramétricos, descritos no domínio da frequência, os mais utilizados são a função de autocovariância (ou autocorrelação) e sua transformada de Fourier ou espectro de frequência e a análise de *Wavelets*. Conforme destacado por Morettin e Tolo (2006), embora seja muito comum trabalhar com modelos paramétricos, a análise espectral desempenha um papel de grande relevância nos estudos de resposta em frequências.

### **2.1.1 Séries temporais não-estacionárias**

As séries temporais podem ser analisadas a partir da sua estacionariedade. Uma série temporal é dita estacionária quando ela se desenvolve ao longo do tempo, de maneira aleatória, ao redor de uma média constante (MORETTIN; TOLOI, 2006). As variáveis econômicas e de processos industriais costumam apresentar algum tipo de não-estacionariedade. Estas séries necessitam de uma transformação prévia com intuito

de remover ou, pelo menos, mitigar as não-estacionariedades existentes (HAMILTON, 1994).

A transformação mais comum consiste em realizar sucessivas operações de diferenciação na série original, até que seja obtida uma série estacionária. Neste sentido, dada uma série temporal  $Z_t$ , com  $N$  amostras, a primeira diferença é definida pela Equação (1):

$$\Delta Z_t = Z_t - Z_{t-1} \quad (1)$$

em que  $\Delta$  representa a operação de diferença, para os instantes de tempo  $t = 2, \dots, N$ . Assim, a segunda diferença será dada pela Equação (2):

$$\Delta^2 Z_t = \Delta(\Delta Z_t) = \Delta(Z_t - Z_{t-1}) = Z_t - 2Z_{t-1} + Z_{t-2} \quad (2)$$

para os instantes de tempo  $t = 3, \dots, N$ . Portanto, de modo geral, a  $d$ -ésima diferença de  $Z_t$  é descrita pela Equação (3),

$$\Delta^d Z_t = \Delta(\Delta^{d-1} Z_t) \quad (3)$$

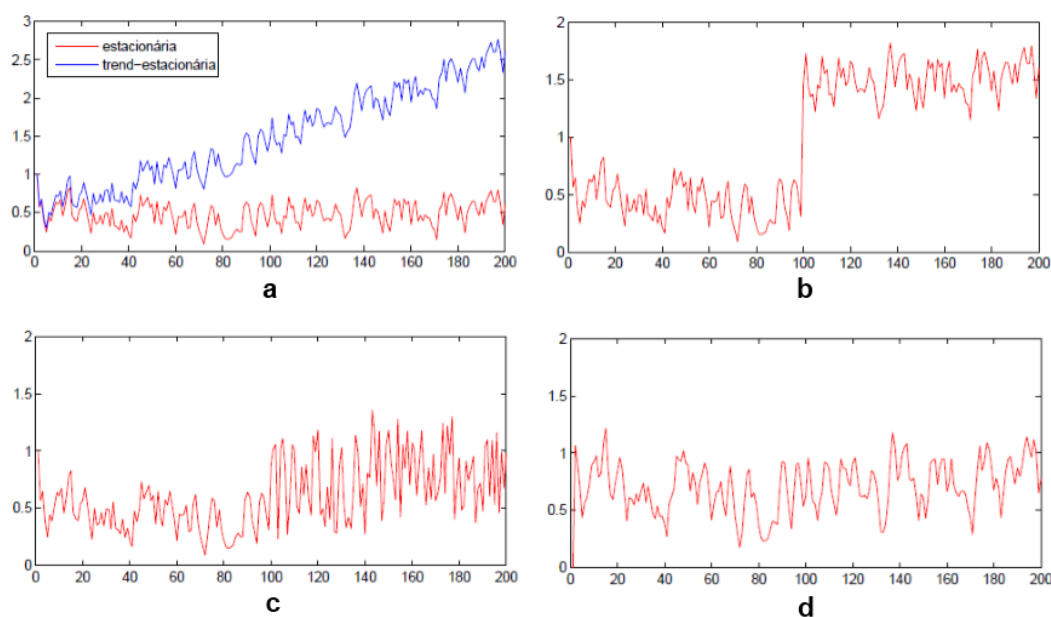
para os instantes de tempo  $t = d + 1, \dots, N$ . Em situações normais, uma ou duas diferenciações já são suficientes para que a série se torne estacionária (MORETTIN; TOLOI, 2006). No Anexo A são apresentadas duas das principais classes de modelos paramétricos utilizados na modelagem de séries temporais não-estacionárias.

#### 2.1.1.1 Principais tipos de não-estacionariedade

A Figura 1 apresenta alguns dos principais tipos de séries temporais não-estacionárias. A Figura 1 (a) ilustra, em azul, um processo *trend*-estacionário, pois a série flutua ao redor de uma tendência determinística descrita por uma função afim. No entanto, em alguns casos esta tendência é tão sistemática que é possível analisar o desvio da tendência como uma variável estacionária, produzindo a série representada em vermelho, permitindo-se a aplicação do processo usual de análise de séries estacionárias. A Figura 1 (b) também apresenta outro tipo de não-estacionariedade, representada por séries que possuem variações na média a partir de algum ponto no tempo. Estas mudanças de nível podem ser associadas às transições entre pontos de operação de uma planta, por exemplo (HAMILTON, 1994). Na Figura 1 (c) também é possível visualizar as séries com mudança de variância e, assim como nas séries com mudança de nível, estas mudanças representam regimes de operação diferentes, em que um regime é mais volátil que o outro. A Figura 1 (d)

mostra um exemplo de série com raízes unitárias no polinômio característico, gerando um modelo não-estacionário denominado de caminhada aleatória (*random walk*). Nota-se que a caminhada aleatória tende a oscilar livremente (HAMILTON, 1994; MORETTIN; TOLOI, 2006).

Figura 1 – Exemplos de séries temporais não-estacionárias com 200 amostras no tempo: a) Processo estacionário e *trend*-estacionário; b) Processo com mudança de nível; c) Processo com mudança de variância; d) Processo com raízes unitárias (caminhada aleatória).



Fonte: Adaptado de Gomes (2016, p. 31).

### 2.1.2 Testes de Estacionariedade

Quando a média e a variância de uma determinada série não se apresentam constantes ao longo do tempo é considerada a existência de pelo menos uma raiz unitária. A partir da década de 70, foram elaborados diversos testes de raízes unitária, como o Teste de Dickey-Fuller (DF), o Teste de Dickey-Fuller Aumentado (*Augmented Dickey-Fuller* – ADF) e o teste não-paramétrico de Phillips-Perron (PP). Estes testes verificam a hipótese nula ( $H_0$ ) de que existe uma raiz unitária, contra uma hipótese alternativa ( $H_1$ ) de que a série é estacionária e, portanto, não há raiz unitária (BUENO, 2011; HAMILTON, 1994; MORETTIN; TOLOI, 2006).

Contudo, um dos principais problemas dos testes de raiz unitária é o seu baixo poder discriminatório, particularmente por não conseguirem distinguir um processo estacionário de outro quase estacionário, apresentando, por exemplo, a presença de



um componente de média móvel muito próximo da região de fronteira não estacionária. Desta forma, os resultados destes testes são estritamente influenciados não só pela presença ou não de uma constante e/ou tendência, como também pelo próprio tamanho da amostra utilizada (BUENO, 2011; KWIATKOWSKI *et al.*, 1992). Por esta razão, partindo da premissa contrária aos tradicionais testes de raiz unitária, foram desenvolvidos testes alternativos os quais objetivam-se testar a estacionariedade da série. Um deles é o teste KPSS (Kwiatkowski, Phillips, Schmidt, and Shin), que testa a hipótese nula de estacionariedade da série, sendo capaz de distinguir a raiz unitária de séries cujos dados não são suficientemente conclusivos (BUENO, 2011).

### 2.1.3 Testes de Linearidade

Desde a contribuição de Box e Jenkins (1976), a identificação e estimativa de modelos lineares tornaram-se ferramentas estatísticas tradicionais para análise de séries temporais. No entanto, os modelos lineares podem deixar de capturar por completo o comportamento dinâmico de séries temporais reais, uma vez que estas são muitas vezes caracterizadas por componentes não-lineares significativas (CORDUAS, 1994; SAXEN; GAO, C.; GAO, Z., 2013; TUNÇKAYA; KÖKLÜKAYA, 2016). Por esse motivo, é importante que durante a análise preliminar sejam realizadas verificações quanto à linearidade do processo a ser modelado (CORDUAS, 1994; GOMES, 2016).

O teste de Lilliefors (LILLIEFORS, 1967) é um teste de linearidade, baseado no teste de Kolmogorov-Smirnov (KOLMOGOROV, 1933; SMIRNOV, 1948), usado para verificação de normalidade de um conjunto de dados. Ele é usado para verificar a hipótese nula de que os dados apresentam distribuição normal  $N(\mu, \sigma^2)$ , sem a necessidade de especificação da função de distribuição proposta em  $H_0$ . Este teste se distingue do teste de Kolmogorov-Smirnov pois ele considera que a média e a variância da população são desconhecidas e, por isso, as mesmas são estimadas baseando-se no pacote de dados testado (GIBBONS; CHAKRABORTI, 2003).

A segunda parte do teste de Lilliefors consiste do cálculo da máxima discrepância (desvio máximo) calculada entre a distribuição de frequências acumuladas que deveria ocorrer sob a suposição de  $H_0$  com a média e variância estimadas e a função de distribuição empírica. A partir de então, é avaliado se o desvio máximo encontrado é estatisticamente significativo, implicando na rejeição de  $H_0$  (GIBBONS;

CHAKRABORTI, 2003; KOLMOGOROV, 1933; LILLIEFORS, 1967; SMIRNOV, 1948).

O teste de Kruskal-Wallis (KRUSKAL; WALLIS, 1952) é um método não-paramétrico utilizado para decidir se dois ou mais conjuntos de amostras provêm da mesma distribuição, pela análise das variâncias. Este teste analisa se existem diferenças efetivas entre as distribuições ou se estas diferenças representam apenas variações causais. O teste é baseado na hipótese nula de que todas as populações possuem funções de distribuição iguais contra a hipótese alternativa de que ao menos duas das populações possuem funções de distribuição diferentes (GIBBONS; CHAKRABORTI, 2003).

Caso sejam constatadas a presença de características não-lineares no comportamento das séries temporais, há necessidade de se investigar o desempenho de técnicas capazes de lidar diretamente com as mesmas, como é o caso das RNA, devido às unidades de processamento não-lineares que podem ser usadas nesta construção (HAYKIN, 2001; SAXEN; GAO, C.; GAO, Z., 2013).

Dorffner (1996), ao realizar uma revisão das RNA no processamento de séries temporais, destaca a sua superioridade frente aos modelos estatísticos tradicionais em relação à capacidade de aproximação de funções não-lineares. A modelagem de séries temporais empregando as redes neurais será abordada no Capítulo 3.

## 2.2 A TRANSFORMADA WAVELET

As bases matemáticas formuladas por Joseph Fourier no início do século XIX permitiram mapear um sinal no domínio do tempo para o domínio da frequência, através da soma de infinitas componentes senoidais deste sinal.

Assim, a Transformada de Fourier (*Fourier Transform* – FT) de um sinal  $x(t)$ , denotada por  $\mathcal{F}[x(t)]$ , é definida através da Equação (4):

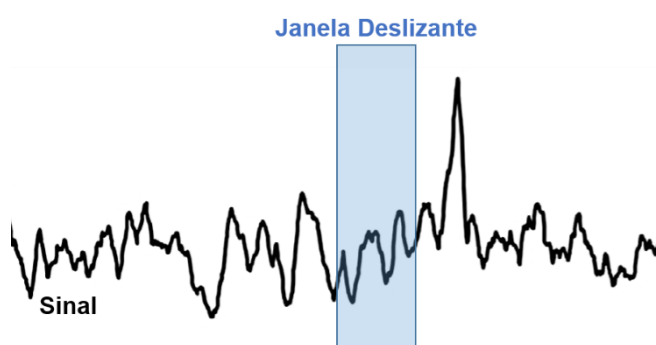
$$X(f_s) = \mathcal{F}[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j2\pi f_s t} . dt \quad (4)$$

em que  $t$  é usado para representar o tempo e  $f_s$  a frequência do sinal. Embora esta transformada apresente máxima resolução em frequência, ela não apresenta resolução no tempo. Desta forma, quando um sinal é representado pela FT, assume-se que as suas componentes de frequência são invariantes no tempo e, por isso, não se tem informações sobre quais momentos estas frequências ocorrem (OLIVEIRA,

2007b). Partindo desta premissa, com o intuito de possibilitar o tratamento de sinais não-estacionários, Gabor (1946) sugeriu uma modificação na equação da FT, conhecida como Transformada de Fourier de Tempo-Curto (*Short-Time Fourier Transform* – STFT).

A STFT utiliza uma janela deslizante (ver Figura 2) de tamanho fixo que percorre o sinal, analisando uma pequena seção temporal do mesmo, dentro da qual o sinal permanece aproximadamente estacionário.

Figura 2 – O uso de janelas deslizantes na análise local do tempo.



Fonte: O próprio autor.

Conforme descreve a Equação (5), STFT trata-se de uma representação bidimensional, por meio do tempo e frequência:

$$STFT(\tau, f_s) = \int_{-\infty}^{\infty} x(t)g(t - \tau)e^{-j2\pi f_s t}.dt \quad (5)$$

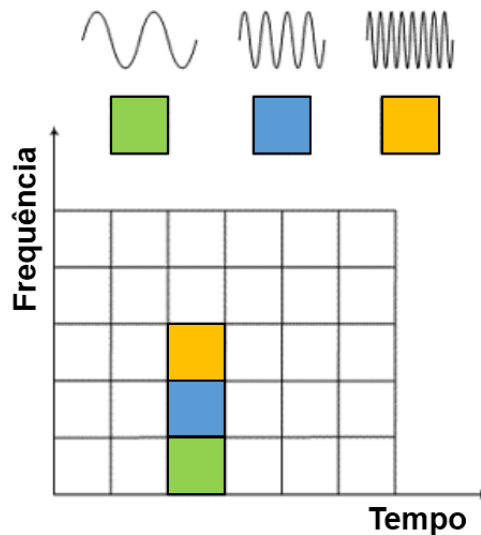
onde o sinal  $g(t - \tau)$  é a função responsável pelo janelamento centrado em  $\tau$ , em geral, dado por uma gaussiana.

Apesar da STFT ser capaz de fornecer a informação de quando uma determinada frequência ocorreu no sinal, esta informação tem abrangência limitada pelo tamanho da janela, que será o mesmo para todas as frequências (POLIKAR, 1996). Grossmann e Morlet (1984) ao analisar sinais geofísicos destinados a exploração do petróleo, perceberam que a análise feita através das STFT ainda é ineficaz em situações em que o sinal apresenta características não-estacionárias, isto é, sinais cujo espectro de frequências varia ao longo do tempo.

Em geral, a dinâmica dos sinais de altas frequências pode ser compreendida após a análise de um curto intervalo de tempo, em contraste com os sinais de baixas frequências, que devem ser observados em um intervalo relativamente maior.

Conforme ilustra a Figura 3, uma vez que, fixada uma janela para a STFT, a resolução no tempo  $\Delta t$  e a resolução em frequência  $\Delta f_s$  permanecem fixas em todo plano tempo-frequência. Sendo assim, o sinal não poderá ser examinado em alta resolução em relação ao tempo e à frequência simultaneamente (RIOUL; VETTERLI, 1991).

Figura 3 – Diagrama do plano tempo-frequência pela análise da STFT.



Fonte: O próprio autor.

Assim, a Transformada *Wavelet* (*Wavelet Transform* – WT) foi desenvolvida como uma abordagem alternativa a STFT para superar o seu problema de resolução. Esta ferramenta permite que um sinal seja decomposto em diferentes componentes de frequências, possibilitando assim, o estudo de cada componente separadamente em sua escala correspondente. A Equação (6) mostra a relação matemática referente à WT:

$$CWT(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}(t) dt \quad (6)$$

em que  $a$  é o parâmetro de escala,  $b$  é o parâmetro de translação,  $x(t)$  é o sinal no tempo e  $\psi_{a,b}(t)$  é a função *wavelet* base,  $\psi(t)$ , escalada e deslocada conforme a Equação (7):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (7)$$

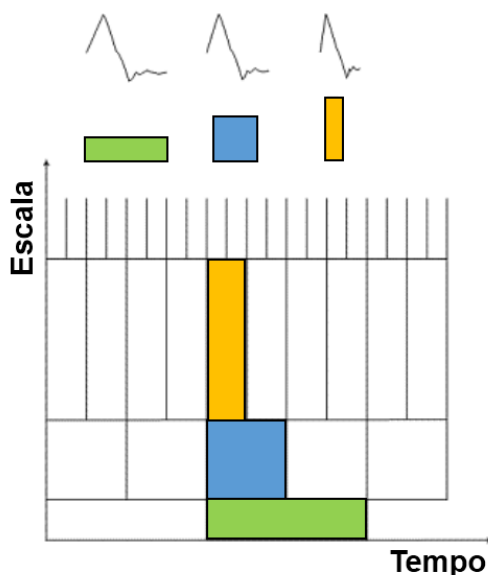
É importante destacar que na WT a noção de escala é introduzida como uma alternativa para a frequência, visto que o termo frequência é reservado a FT. Assim,

o sinal é então representado em um plano tempo-escala (OLIVEIRA, 2007a; RIOUL; VETTERLI, 1991).

A WT é uma ferramenta poderosa para o processamento de sinal, podendo ser aplicada não só na separação de componentes de frequência no sinal, como também na compressão de dados, remoção de ruídos e processamento de imagem. Nesta transformada, as resoluções do tempo e de escala são variadas de forma a obter uma Análise Multiresolução (*MultiResolution Analysis* – MRA), capaz de analisar o sinal a partir de diferentes componentes de frequências em diferentes resoluções temporais (MALLAT, 1989; MISITI *et al.*, 2013).

De acordo com Mallat (1989), a MRA é uma ferramenta útil, uma vez que, boa parte dos sinais encontrados nas aplicações práticas apresentam componentes de alta frequência por curtas durações de tempo e componentes de baixa frequência por longa duração de tempo. Através da análise por *wavelet*, são aplicadas janelas temporais pequenas, onde se deseja maiores informações de alta frequência, e janelas temporais maiores, onde se deseja informações de baixa frequência, conforme evidencia a Figura 4.

Figura 4 – Diagrama do plano tempo-frequência pela análise da WT.



Fonte: O próprio autor.

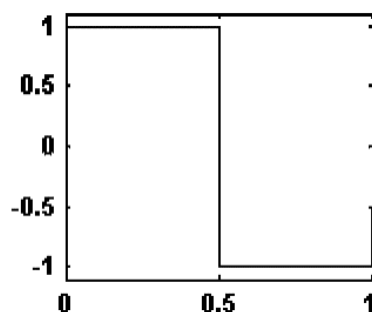
Esta particularidade é importante na análise de sinais, uma vez que, em geral, são os componentes de baixa frequência que caracterizam o comportamento de um

determinado sinal, enquanto que os componentes de alta frequência são os responsáveis por fornecer os detalhes do mesmo (RIOUL; VETTERLI, 1991).

### 2.2.1 A Base *Wavelet* de Haar

A função Haar é a primeira e a mais simples das *wavelets* (Figura 5). Ela foi introduzida em 1910, por Alfred Haar (HAAR, 1910) e trata-se de uma *wavelet* descontínua, que utiliza pulsos quadrados para aproximar a função original, conforme Equação (8). A sua utilização é adequada para sinais que possuem mudanças abruptas de valores ou discontinuidades no decorrer do tempo. Por simplicidade, a sua representação matemática considera um sinal constante por partes (MISITI *et al.*, 2013).

Figura 5 – Exemplo de uma *wavelet* de Haar.



Fonte: Misiti *et al.* (2013, p. 94).

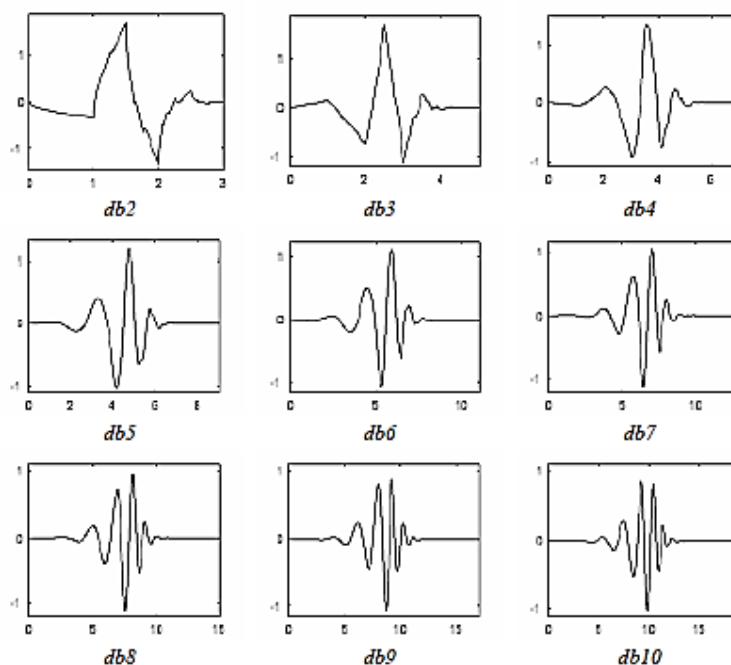
$$\psi(t) = \begin{cases} 1 & \text{para } 0 \leq t < 1/2 \\ -1 & \text{para } 1/2 \leq t < 1 \\ 0 & \text{caso contrário} \end{cases} \quad (8)$$

### 2.2.2 Bases *Wavelets* de Daubechies

A partir da família de *wavelets* formulada por Ingrid Daubechies (DAUBECHIES, 1988), a análise discreta de *wavelet* tornou-se possível. As *wavelets* de Daubechies são uma família de *wavelets* perfeitamente compactadas no tempo, embora que no domínio da frequência apresente um alto grau de superposição espectral entre as escalas. A característica mais relevante desta família é que por serem ortogonais, um erro no sinal de entrada não cresce com a transformação e a estabilidade numérica computacional é assegurada. Os nomes da família Daubechies são escritos como  $dbN_o$ , onde  $N_o$  é a sua ordem. As *wavelets* desta família para as ordens de 2 a 10 são

apresentadas na Figura 6. A *wavelet* Haar pode ser definida como uma Daubechies de ordem 1, isto é, *db1* (MISITI *et al.*, 2013; OLIVEIRA, 2007b).

Figura 6 – Representação das funções da família Daubechies de *db2* até *db10*.

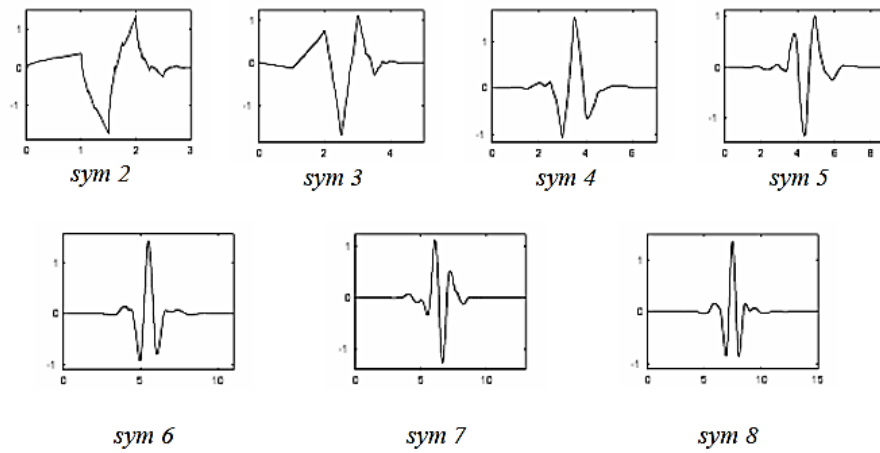


Fonte: Misiti *et al.* (2013, p. 95).

### 2.2.3 Bases *Wavelets* de Symlets e Coiflets

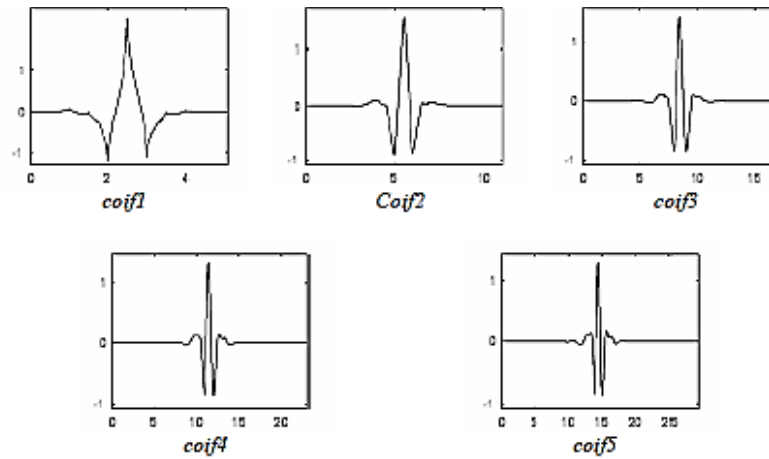
Essas duas famílias são derivadas das *wavelets* de Daubechies, construídas pela própria Daubechies a pedido de Ronald Coifman em 1989 (DAUBECHIES, 1988, 1992). Ela propôs modificações na construção das  $dbN_o$ , de forma que a simetria fosse aumentada sem comprometimento de sua simplicidade. Em geral, as Coiflets tendem a ser mais simétricas que as Symlets. Assim como a família Daubechies, as famílias Symlets e Coiflets podem ser representadas, respectivamente, como  $symN_o$  e  $coifN_o$  (MISITI *et al.*, 2013; OLIVEIRA, 2007b). As *wavelets* Symlets de ordem 2 a 8 são apresentadas pela Figura 7, seguida pela Figura 8, a qual apresenta as Coiflets de ordem 1 a 5.

Figura 7 – Representação das funções da família Symlets de *sym2* até *sym8*.



Fonte: Misiti *et al.* (2013, p. 97).

Figura 8 – Representação das funções da família Coiflets de *coif1* até *coif5*.



Fonte: Misiti *et al.* (2013, p. 98).

#### 2.2.4 Transformada Wavelet Discreta

Na Transformada Wavelet Discreta (*Discrete Wavelet Transform – DWT*), a família de wavelets  $\psi_{j,r}(t)$  é dada pela Equação (9):

$$\psi_{j,r}(t) = \frac{1}{\sqrt{|a_0^j|}} \psi\left(\frac{t - rb_0a_0^j}{a_0^j}\right) \quad (9)$$

em que  $j$  e  $r$  são números inteiros, correspondentes aos parâmetros de dilatação e translação, respectivamente,  $a_0 > 1$  é um parâmetro de dilatação fixo, que reflete a



variação da dilatação, e  $b_0$  é o fator de translação (passo da translação) o qual depende do fator de dilatação (POLIKAR, 1996).

Em geral, escolhe-se  $a_0 = 2$  e  $b_0 = 1$  para que sejam obtidas amostragens de frequência e de tempo diáticas, resultando na Equação (10):

$$\psi_{j,r}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j r}{2^j}\right) = 2^{-\frac{j}{2}} \psi(2^{-j} t - r) \quad (10)$$

Desta forma, obtém-se uma escala de dilatação como uma potência de dois ( $a_0 = 2^j$ ) e passos de translação proporcionais à esta escala ( $b_0 = 2^j r$ ).

De acordo com Mallat (1989), a DWT pode ser implementada, baseando-se na MRA, a qual permite analisar o sinal em múltiplas bandas de frequência, a partir de filtros passa-baixas (*Low-Pass Filter* – LPF) e filtros passa-altas (*High-Pass Filter* – HPF), de tal modo que a cada iteração o espectro seja dividido ao meio. O algoritmo fundamenta-se na ideia de que, a cada iteração, sejam criadas duas séries de coeficientes:  $cD$  e  $cA$ , respectivamente, coeficientes de detalhe e aproximação.

O filtro passa-altas está relacionado à função *wavelet*  $\psi(t)$  e é o responsável por produzir os coeficientes de detalhes da decomposição (ver Equação (11)). Na MRA, há uma função adicional, denominada função escalonamento  $\phi(t)$ , que também pode ser dilatada e transladada tal como a função *wavelet*. A função  $\phi(t)$  está relacionada ao filtro passa-baixas e associada aos coeficientes de aproximação da decomposição, de acordo com a Equação (12) (MALLAT, 2008; PERCIVAL; WALDEN, 2000; OLIVEIRA, 2007b).

$$d_{j,r} = \int_{-\infty}^{\infty} x(t) \psi_{j,r}(t) dt \quad (11)$$

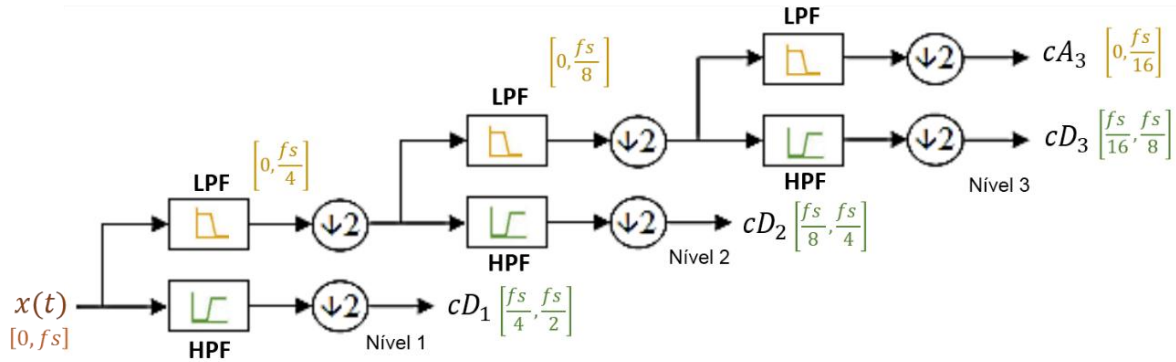
$$a_{J,r} = \int_{-\infty}^{\infty} x(t) \phi_{J,r}(t) dt \quad (12)$$

em que  $a_{J,r}$  são os componentes de aproximação na escala máxima,  $J$ , e  $d_{j,r}$  são os coeficientes de detalhe em todas as escalas de 1 a  $J$ . O valor  $J$  representa o número de níveis e  $r$  varia entre 1 e o número máximo de coeficientes da componente correspondente.

Os filtros digitais utilizados nesta análise são chamados de Filtros de Quadratura (*Quadrature Filter* – QF), responsáveis por filtrar e decimar o sinal. A decimação (*downsampling*) é um procedimento de descarte espaçado de amostras de um sinal

discreto, diminuindo a frequência de amostragem e que, conseqüentemente, aumenta-se o tamanho do período de amostragem. Neste sentido, decimar o sinal por um fator 2, significa que a cada dois pontos de dados, um será eliminado, obtendo-se um sinal com metade do comprimento inicial (OPPENHEIM; SCHAFER; BUCK, 1999). Portanto, conforme ilustra a Figura 9, a implementação da DWT pode ser entendida como um processo de decomposição iterativo com sucessivas decomposições nos coeficientes de aproximação  $cA$ , através de uma estrutura de banco de filtros. Após o sinal  $x(t)$ , passar pelos filtros LPF e HPF, sofre decimação e dá origem às séries de coeficientes  $cA$  e  $cD$ , cada um com metade da dimensão do sinal original (MISITI *et al.*, 2017; POLIKAR, 1996).

Figura 9 – Processo de decomposição do sinal via DWT.



Fonte: O próprio autor.

A DWT pode então ser formulada considerando uma decomposição da série original em subséries. Assim, um sinal  $x(t)$  pode ser representado conforme a Equação (13):

$$x(t) = cA_J(t) + cD_J(t) + cD_{J-1}(t) + \dots + cD_1(t) = cA_J(t) + \sum_{j=1}^J cD_j(t) \quad (13)$$

onde  $cA_j(t)$  e  $cD_j(t)$  são as séries de componentes de aproximação e detalhe, respectivamente, dadas pelas Equações (14) e (15):

$$cA_j(t) = \sum_r a_{j,r} \phi_{j,r}(t) \quad (14)$$

$$cD_j(t) = \sum_r d_{j,r} \psi_{j,r}(t), \quad j = 1, \dots, J \quad (15)$$

Uma vez que, devido à decimação, a frequência máxima do sinal amostrado vale metade do seu valor, após o sinal passar pelo filtro passa-baixas de meia banda,

metade das amostras serão eliminadas. Deve-se frisar, no entanto, que a reconstituição do sinal é garantida pelo Teorema de Nyquist, que estabelece que os sinais devem ser amostrados com uma frequência superior ao dobro de sua frequência máxima e, portanto, se um sinal é amostrado a uma frequência  $f_s$ , a análise espectral só deve considerar as frequências inferiores a  $f_s/2$  (OLIVEIRA, 2017b; POLIKAR, 1996). Portanto, a principal vantagem da decimação é que são mantidas apenas as informações suficientes que permitam a reconstrução exata dos dados de entrada (RENAUD; MURTAGH; STARCK, 2002).

#### 2.2.4.1 A *Wavelet* Packet Discreta

A cada iteração da DWT, apenas os coeficientes de aproximação são divididos em novos componentes  $cA$  e  $cD$ . Em contraste com a DWT, a Transformada *Wavelet* Packet Discreta (*Discrete Wavelet Packet Transform* – DWPT), permite que os elementos  $cD$  também sejam divididos em novos componentes  $cA$  e  $cD$  (MALLAT, 1998), permitindo assim fragmentar ainda mais a representação do sinal (PESQUET; KRIM; CARFANTAN, 1996).

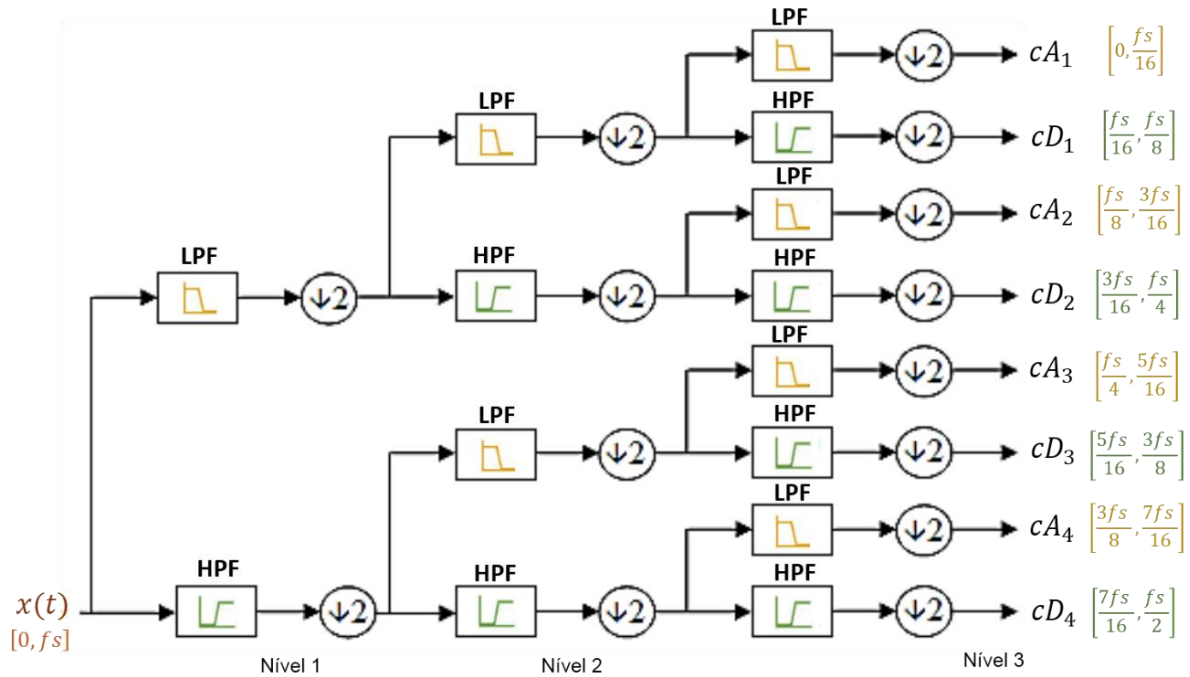
Nas transformações da DWT, os coeficientes de aproximação identificam a faixa de frequência  $[0, f_s/(2^{j+1})]$ , enquanto que os coeficientes de detalhe no nível  $j$  descrevem a banda de frequência  $[f_s/(2^{j+1}), f_s/(2^j)]$ . Por outro lado, a DWPT permite que, em cada nível  $j$ , a banda de frequência seja dividida em  $2^j$  partições de comprimentos iguais (ver Figura 10). Desta forma, o sinal passa a ser representado em bandas de frequências com diferentes resoluções (MISITI *et al.*, 2013; PERCIVAL; WALDEN, 2000).

Em termos matemáticos, a DWPT pode ser representada conforme a Equação (16):

$$x(t) = \sum_{j=1}^J cA_j(t) + \sum_{j=1}^J cD_j(t) \quad (16)$$

No entanto, nas abordagens DWT e DWPT, o comprimento do sinal é limitado a um múltiplo inteiro de uma potência de dois. Neste sentido, sejam  $L$  o nível de decomposição máximo,  $\text{int}(\cdot)$  a função que retorna o número inteiro mais próximo e  $N$  o comprimento do sinal de entrada, o nível de decomposição máximo é definido pela Equação (17) (MISITI *et al.*, 2017).

Figura 10 – Processo de decomposição do sinal via DWPT.



Fonte: O próprio autor.

$$L = \text{int}(\log_2 N) \quad (17)$$

Portanto, uma limitação apresentada pelas abordagens DWT e DWPT é que para se decompor algum sinal em um nível  $L$ , é necessário que o tamanho da amostra seja um múltiplo inteiro de  $2^L$  (PERCIVAL; MOFJELD, 1997). Além disto, outra limitação apresentada por estas abordagens é a sua variância no tempo. Esta propriedade é não é desejável em aplicações como detecção ou estimação de parâmetros de sinais com constante atualização dos dados. Isto implica que, caso fossem excluídos os primeiros valores da série de entrada, então as subséries resultantes da transformação não seriam idênticas às subséries obtidas antes da exclusão (PESQUET; KRIM; CARFANTAN, 1996; RENAUD; MURTAGH; STARCK, 2002).

Este problema, no entanto, pode ser resolvido por meio de Transformadas *Wavelets* não-decimadas e invariantes no tempo, tais como a *Maximal Overlap Discrete Wavelet Transform* – MODWT) e a sua versão Packet (*Maximal Overlap Discrete Wavelet Packet Transform* – MODWPT). Por não apresentar decimação, este tipo de transformação aplicado à uma série temporal de qualquer comprimento  $N$ , resulta em uma escala de resolução também de comprimento  $N$  para cada um dos níveis de resolução considerados. É fácil, nestas circunstâncias, relacionar informações em

cada escala de resolução para o mesmo ponto de tempo (RENAUD; MURTAGH; STARCK, 2002).

### 2.3 TRANSFORMADA WAVELET APLICADA NA ANÁLISE DE SÉRIES TEMPORAIS

A MODWT e a MODWPT são transformadas não-ortogonais e redundantes, capazes de decompor uma série temporal em subséries, de modo que o sinal seja representado em bandas de frequência com diferentes resoluções (WICKERHAUSER, 1994; PERCIVAL; GUTTORP, 1994). Elas têm sido empregadas em muitos campos da previsão de séries temporais não-estacionárias, através da decomposição dos sinais em subséries construtivas mais simples (ADAMOWSKI, 2008; ALVES *et al.*, 2017; DUNNE; GHOSH, 2013; ISMAIL; YUNUS, 2012; LIU *et al.*, 2013; MURTAGH; STARCK; RENAUD, 2004; RENAUD; MURTAGH; STARCK, 2002; WALDEN; CRISTAN, 1998; WALDEN, 2001; WEI; ZUO; SONG, 2012). Estas subséries são a representação do sinal em bandas de frequências com diferentes resoluções e, por isso, podem ser utilizadas como variáveis de entrada no modelo de previsão (DUNNE; GHOSH, 2013; LIU *et al.*, 2013; MALLAT, 2008; WALDEN; CRISTAN, 1998).

A principal característica destas transformadas é que elas não apresentam sensibilidade ao ponto de partida assumido para uma série temporal, conseqüentemente, não importa quantas e nem a partir de que momento as amostras serão tomadas para análise (LIU *et al.*, 2013; PERCIVAL; WALDEN, 2000). A MODWPT, em especial, tem sido utilizada com frequência como ferramenta analítica para representar sinais na análise de séries temporais não-estacionárias, dividindo a variância dos dados pela escala (MISITI *et al.*, 2017).

Assim como a DWPT, para cada nível  $j$  da MODWPT, a banda de frequência é dividida em  $2^j$  partições de comprimentos iguais e, de acordo com Percival e Walden (2000), os elementos dos coeficientes de aproximação e detalhe obtidos no  $j$ -ésimo nível da MODWPT podem ser escritos conforme as Equações (18) e (19):

$$c\tilde{A}_{j,m} = \sum_{l=0}^{N-1} \tilde{g}_{j,l} x_{m-l \bmod N} \quad (18)$$

$$c\tilde{D}_{j,m} = \sum_{l=0}^{N-1} \tilde{h}_{j,l} x_{m-l \bmod N} \quad (19)$$

com  $j = 1, \dots, J$  e  $m = 0, 1, \dots, N - 1$ . O termo  $m - l \bmod N$  presente nas equações permite uma filtragem circular fazendo com que a série temporal seja representada com os mesmos números de coeficientes em cada escala. Assim, se  $w$  for um número inteiro tal que  $0 \leq w \leq N - 1$ , então  $w \bmod N \equiv w$ ; se  $w$  for qualquer outro inteiro, então  $w \bmod N \equiv w + rN$ , onde  $rN$  é o múltiplo inteiro exclusivo de  $N$  tal que  $0 \leq w + rN \leq N - 1$ . Os termos  $c\tilde{A}_{j,m}$  e  $c\tilde{D}_{j,m}$  caracterizam, respectivamente, o  $m$ -ésimo elemento do  $j$ -ésimo nível dos coeficientes de aproximação e detalhe,  $\tilde{g}_{j,l}$  e  $\tilde{h}_{j,l}$  representam os coeficientes dos filtros passa-baixas e passa-altas, respectivamente, obtidos a partir dos correspondentes filtros da DWT seguindo as Equações (20) e (21):

$$\tilde{g}_l = \frac{g_l}{2^{j/2}} \quad (20)$$

$$\tilde{h}_l = \frac{h_l}{2^{j/2}} \quad (21)$$

no qual  $g_l$  e  $h_l$  são, respectivamente, os coeficientes dos filtros escala (passa-baixas) e *wavelet* (passa-altas) da DWT. Os coeficientes  $\tilde{g}_l$  e  $\tilde{h}_l$  devem satisfazer as seguintes propriedades dadas pelas Equações (22) e (23):

$$\sum_{l=0}^{L_{cf}-1} \tilde{g}_l = 1, \quad \sum_{l=0}^{L_{cf}-1} \tilde{g}_l^2 = \frac{1}{2} \quad e \quad \sum_{l=-\infty}^{\infty} \tilde{g}_l \tilde{g}_{l+2N} = 0 \quad (22)$$

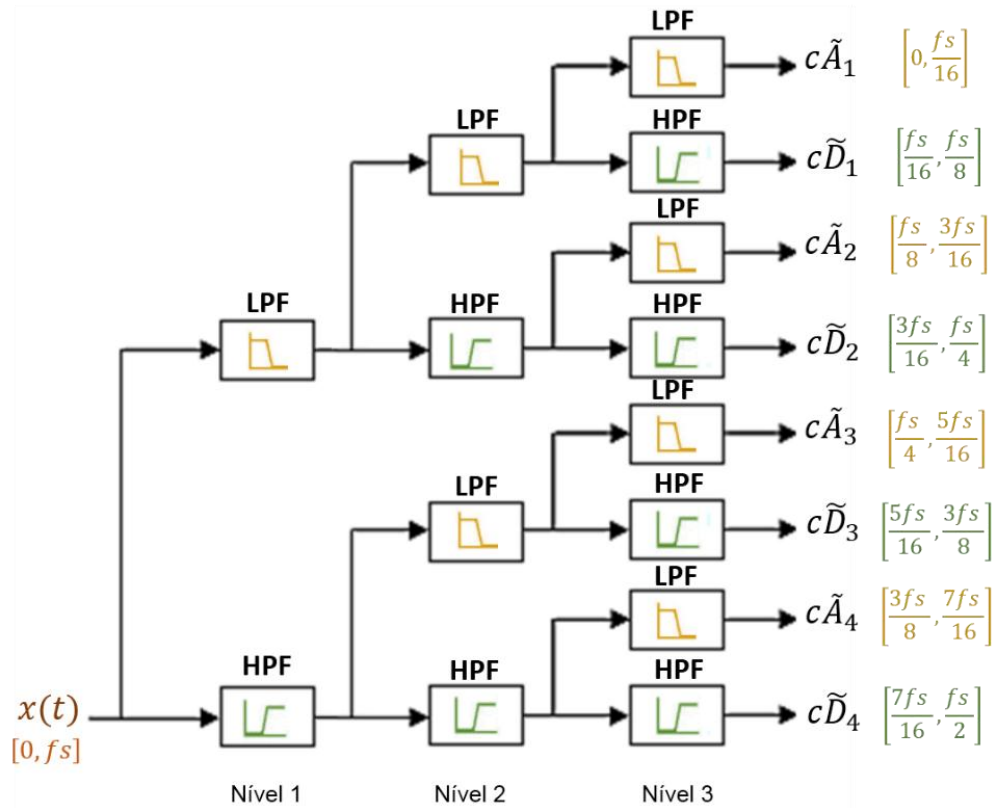
$$\sum_{l=0}^{L_{cf}-1} \tilde{h}_l = 0, \quad \sum_{l=0}^{L_{cf}-1} \tilde{h}_l^2 = \frac{1}{2} \quad e \quad \sum_{l=-\infty}^{\infty} \tilde{h}_l \tilde{h}_{l+2N} = 0 \quad (23)$$

para todo  $n$  inteiro não-nulo e  $l = 0, 1, \dots, L_{cf} - 1$ , na qual  $L_{cf}$  é o número de coeficientes dos filtros escala e *wavelet* (PERCIVAL; WALDEN, 2000).

Um processo de decomposição da MODWPT em três níveis é ilustrado pela Figura 11, em que o sinal  $x(t)$  é decomposto em 8 subséries.

Da mesma maneira como na DWPT, baseando-se na Figura 11, é possível descrever um determinado sinal  $x(t)$  como um somatório das subséries resultantes da decomposição, conforme evidencia a Equação (24):

Figura 11 – Processo de decomposição do sinal via MODWPT.



Fonte: O próprio autor.

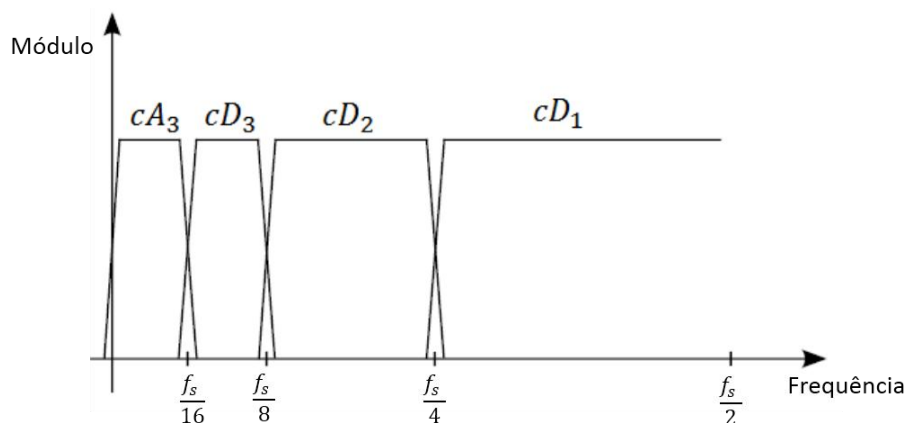
$$x(t) = \sum_{j=1}^J c\tilde{A}_j + \sum_{j=1}^J c\tilde{D}_j \quad (24)$$

Outra característica das subséries resultantes do processo de decomposição via MODWPT é a sua perfeita reconstituição aditiva. Neste sentido, durante o processo de reconstrução, os coeficientes da decomposição passam por filtros passa-baixas e passa-altas reversos de maneira que o sinal original possa ser perfeitamente reconstruído (MALLAT, 2008; MISITI *et al.*, 2017).

A MODWPT supera as desvantagens da DWT e da DWPT pois não só apresenta largura de banda de frequência uniforme, como também, ela é invariante no tempo. Além disto, elas não apresentam efeito de borda nas extremidades da série temporal, bem como permitem a adição de amostras na série sem a necessidade de recalculá-la toda a transformada novamente. Essas propriedades são muito desejáveis ao lidar com séries temporais (PERCIVAL; WALDEN, 2000; WICKERHAUSER, 1994).

Nas abordagens MODWT e MODWPT, as componentes não apresentam ortogonalidade e, por isso, há uma leve sobreposição das bandas de frequências (ver Figura 12) e isso ocorre como efeito colateral para se resolver o problema do efeito de borda e permitir o seu uso em séries temporais. Todavia, essa sobreposição não chega a afetar de forma negativa a sua aplicação (MALLAT, 2008; PERCIVAL; WALDEN, 2000; RENAUD; MURTAGH; STARCK, 2002; WALDEN; CRISTAN, 1998).

Figura 12 – Sobreposição das bandas de frequência de um processo MODWT.



Fonte: O próprio autor.



### 3 REDES NEURAIS ARTIFICIAIS

Inspirada na arquitetura e aprendizagem do cérebro humano, as RNA apresentam um comportamento auto-adaptativo, sem a necessidade de saber muitas informações *a priori* em relação ao problema de estudo, sendo bastante favoráveis na modelagem de processos não-lineares que possuem características complexas e/ou desconhecidas, em contrapartida com os métodos determinísticos tradicionais (BISHOP, 1995; HAYKIN, 2001; ZURADA, 1992).

Nos mais recentes estudos relativos à previsão de séries temporais não-lineares, as redes neurais têm sido aplicadas de maneira enfática, gerando resultados bem superiores aos tradicionais modelos estatísticos já utilizados (BAG, 2007; BULSARI, SAXEN, H.; SAXEN, B., 1992; CHEN, 2001; CHEN; WANG; HAN, 2010; DHOND; GUPTA; VADHAVKAR, 2000; DHOND, 2000; GE, 1999; NURKKALA; PETTERSSON; SAXEN, 2011; SAXEN; PETTERSSON, 2007; SAXÉN; GAO, C.; GAO, Z., 2013; SINGH; SRIDHAR; DEO, 1996; TUNÇKAYA; KÖKLÜKAYA, 2016; WANG *et al.*, 2015).

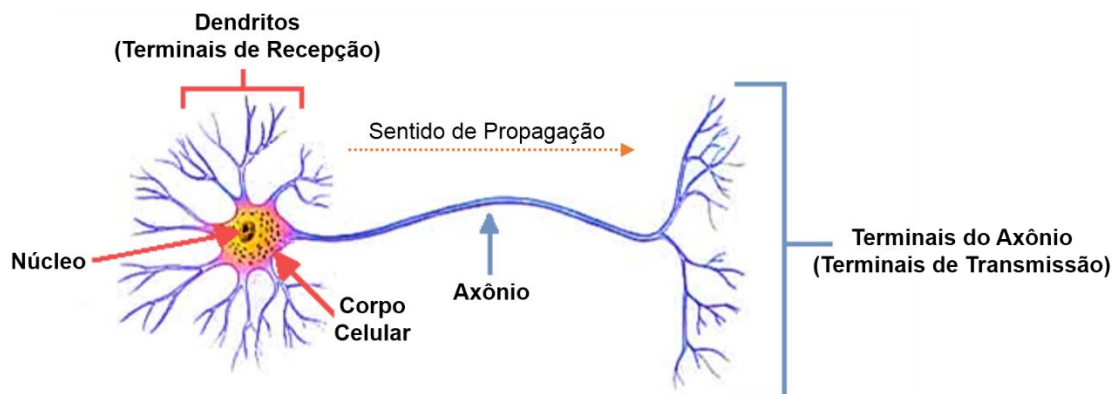
As RNA funcionam como processadores constituídos de unidades de processamento simples, denominados neurônios, que possuem capacidade de armazenar o conhecimento obtido através de experiências e adaptar-se perante às alterações do ambiente, através de um processo de aprendizagem. Devido à sua estrutura flexível, as RNA possuem a capacidade de aproximar funções complexas e, por conseguinte, são capazes de aprender padrões lineares e não-lineares e prevê-los com precisão nos casos em que as técnicas tradicionais falham. Sua natureza paralela a torna menos susceptível às falhas, de forma que, a perda de neurônios locais pouco afeta o comportamento geral da rede (HAYKIN, 2001).

#### 3.1 O MODELO DE NEURÔNIO

O cérebro humano é composto por bilhões de neurônios interconectados e, cada neurônio é composto basicamente por três seções: o corpo celular, os dendritos e o axônio, cada um com funções específicas (ver Figura 13). Os dendritos recebem sinais de entrada, caracterizados por impulsos (ativações) nervosos oriundos de outros neurônios, e os encaminha até o corpo celular, onde a informação é processada e novos impulsos são gerados. Estes novos sinais são transmitidos aos outros

neurônios através dos terminais do axônio (HAYKIN, 2001; KOSKO, 1992; SARANGAPANI, 2006).

Figura 13 – Componentes do neurônio biológico.



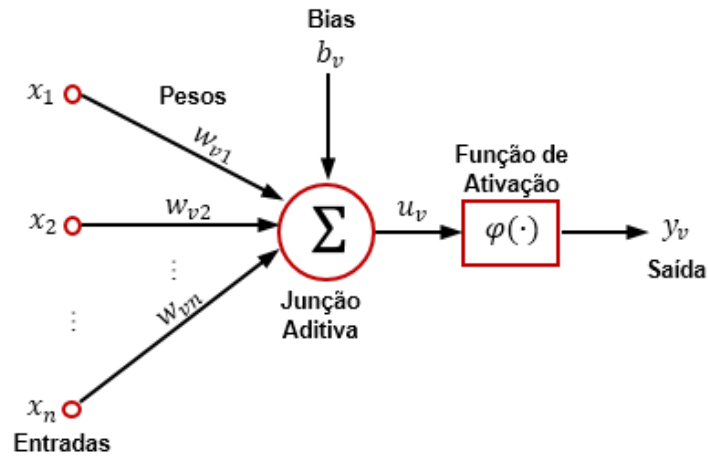
Fonte: O próprio autor.

Os neurônios são conectados entre si por meio da região de sinapse. Esta região é formada pelo contato entre as ramificações do axônio de um neurônio e os dendritos do outro e é através dela que os neurônios se unem funcionalmente, constituindo assim, as redes neurais biológicas. A comunicação neural é dada da seguinte forma: os terminais de recepção recebem ativações excitatórias ou inibitórias dos demais neurônios conectados aos seus dendritos, e caso a combinação destes sinais esteja acima de um determinado limiar de excitação, um impulso elétrico é produzido e propagado para os neurônios seguintes (BRAGA; LUDEMIR; CARVALHO, 2014).

Baseando-se no comportamento funcional dos neurônios biológicos, em 1943 o primeiro modelo matemático de um neurônio biológico foi proposto pelo psicólogo e neurofisiologista Warren McCulloch e o matemático Walter Pitts (MCCULLOCH; PITTS, 1943). Este modelo é formado por conjunto de terminais, que assim como os dendritos, recebem um conjunto de  $n$  entradas  $\{x_1, x_2, \dots, x_n\}$ , equivalentes às ativações dos neurônios anteriores, e uma saída  $y$ , representando o terminal do axônio.

Em analogia ao comportamento das sinapses, os terminais de entrada possuem pesos vinculados  $w_1, w_2, \dots, w_n$ , que multiplicam os respectivos sinais de entrada, determinando o quão representativo é aquela conexão (BRAGA; LUDEMIR; CARVALHO, 2014; HAYKIN, 2001). A Figura 14 ilustra a conexão básica de um Neurônio de McCulloch e Pitts (Modelo MCP), no qual  $\Sigma$  representa a soma ponderada dos sinais de entrada e  $\varphi(\cdot)$  a função de ativação.

Figura 14 – Modelo MCP.



Fonte: O próprio autor.

Em termos matemáticos, um neurônio  $v$  pode ser descrito através do seguinte par de Equações (25) e (26):

$$u_v = \sum_{j=1}^n w_{vj} \cdot x_j \quad (25)$$

$$y_v = \varphi(u_v + b_k) \quad (26)$$

em que  $n$  é o número de sinais de entrada,  $x_j$  é o  $j$ -ésimo sinal de entrada do neurônio,  $w_{vj}$  é o peso sináptico associado ao  $j$ -ésimo sinal de entrada,  $u_v$  é a saída do combinador linear,  $b_v$  é o limiar correspondente ao neurônio  $v$  (*bias*),  $\varphi(\cdot)$  é a função de ativação e  $y_v$  é o sinal de saída do neurônio  $v$  (HAYKIN, 2001).

A função de ativação é a responsável por gerar o sinal de saída do neurônio a partir dos valores do vetor de peso  $\mathbf{w} = (w_1, w_2, \dots, w_j)^T$  e de entrada  $\mathbf{x} = (x_1, x, \dots, x_j)^T$ . Esta função pode ser representada por uma função linear (do tipo linear ou linear por partes) ou não-linear como, por exemplo, as funções sigmóide, tangente hiperbólica, logarítmica e gaussiana.

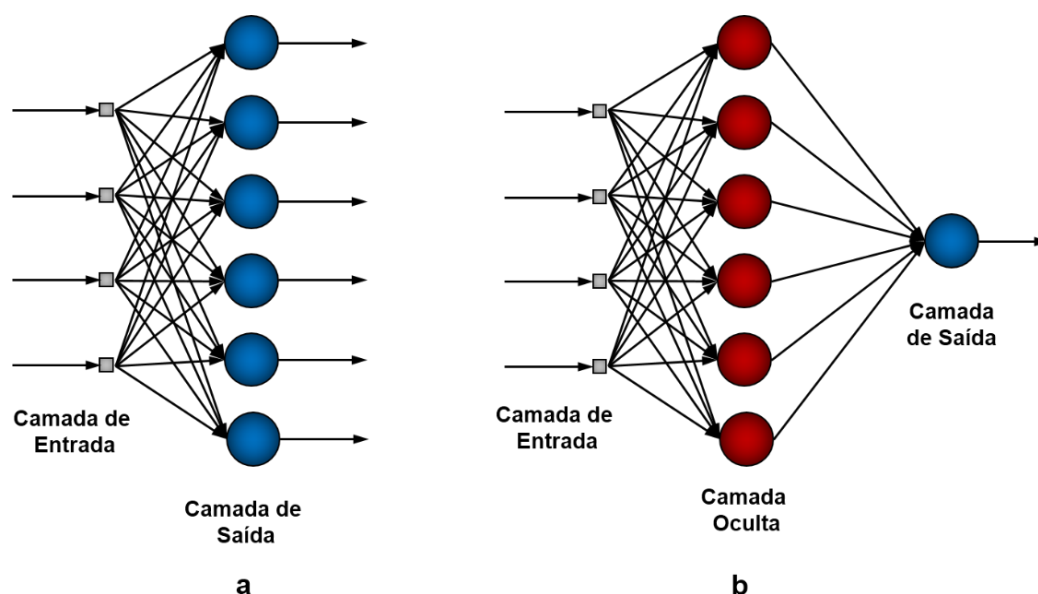
É importante frisar que, independentemente do tipo de função de ativação escolhido, os neurônios individuais podem apresentar limitações. No entanto, um conjunto de neurônios artificiais podem ser estruturados na forma de uma rede neural, capaz de resolver problemas de complexidade elevada (BRAGA; LUDEMIR; CARVALHO, 2014; HAYKIN, 2001; SARANGAPANI, 2006).

### 3.2 PRINCIPAIS TIPOS DE ARQUITETURAS

A forma com que se estabelecem as ligações entre os neurônios define os diferentes tipos de arquiteturas de RNA. As arquiteturas podem ser classificadas em dois grandes grupos: redes com Alimentação para Frente (*Feedforward*) e redes Recorrentes (*Feedback*).

As Redes *Feedforward* permitem apenas o fluxo de sinal unidirecional e seus neurônios são organizados em camadas. Estas redes são consideradas aproximadores universais de funções, sendo capazes de realizar mapeamentos de entrada-saída não-lineares, com alto grau de precisão. Dois exemplos desta topologia são mostrados na Figura 15 (HAYKIN, 2001; HORNIK, 1991; WILAMOWSKI; IRWIN, 2011).

Figura 15 – Redes Neurais Artificiais do tipo *Feedforward*: a) Rede de Camada Única b) Rede com Múltiplas Camadas. Os níveis de bias foram omitidos por conveniência de representação.



Fonte: O próprio autor.

De acordo com Haykin (2001), na sua forma mais simples, conhecida como Rede de Camada Única, os nós da camada de entrada recebem e direcionam os sinais externos de entrada para os neurônios da camada de saída. A designação de camada única refere-se à camada de saída, visto que, na camada de entrada não é realizada nenhum tipo de computação.

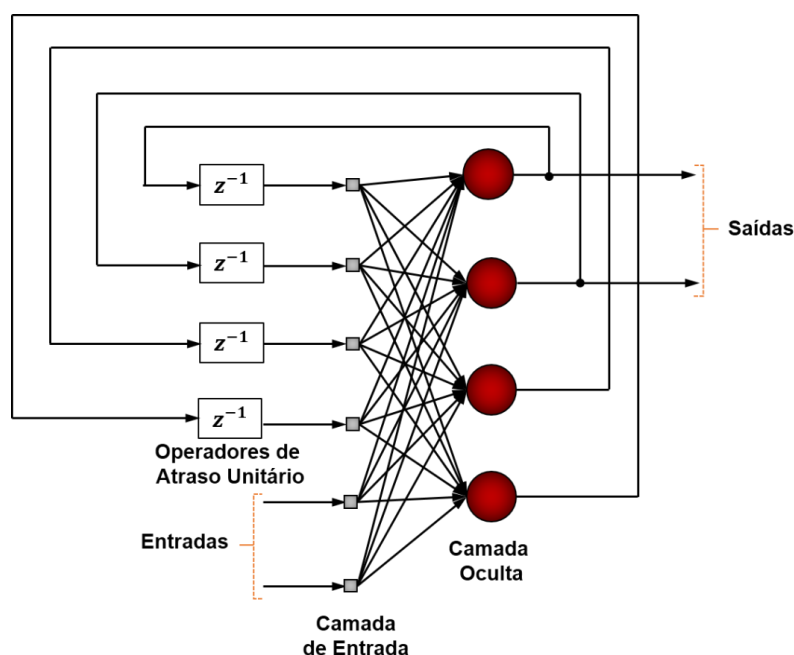
A segunda classe de redes neurais *Feedforward* se distingue pela presença de uma ou mais camadas ocultas, cuja função é receber os sinais da camada anterior,

processá-los e enviá-los para a camada subsequente (HAYKIN, 2001). Cada neurônio da camada oculta apresenta uma função de ativação não-linear continuamente derivável (WILAMOWSKI; IRWIN, 2011), já os neurônios da camada de saída, em geral, apresentam funções de ativação lineares.

Esta arquitetura é também conhecida como Perceptron Multicamada (*Multilayer Perceptron* - MLP). Na prática, são permitidas quaisquer interconexões acíclicas entre neurônios de diferentes camadas, caracterizando o alto grau de conectividade desta rede. Deve-se frisar, ainda, que a adição de camadas intermediárias aumenta a capacidade computacional da rede, no entanto, tende a torná-la mais complexa e requer maior tempo de treinamento (HAYKIN, 2001).

Em contraste com as Redes *Feedforward*, as Redes Recorrentes permitem uma realimentação entre os neurônios de diferentes camadas, fazendo com que a saída não dependa somente das entradas, mas também de informações passadas. Esta estrutura é muito empregada na resolução de problemas de previsão de eventos futuros, pois é capaz de modificar seu aprendizado a partir de informações passadas (BRAGA; LUDEMIR; CARVALHO, 2014). Assim como a MLP, as Redes Recorrentes podem ter ou não a presença de neurônios ocultos. A Figura 16 ilustra um exemplo deste tipo de rede, cujas conexões se originam dos neurônios ocultos como também dos neurônios de saída.

Figura 16 – Exemplo de Rede Recorrente com neurônios ocultos. Os níveis de bias foram omitidos por conveniência de representação.



Fonte: O próprio autor.

A presença dos laços de realimentação tem um forte impacto na capacidade de aprendizagem e desempenho da rede, pois os mesmos são compostos de elementos de atraso unitário, conferindo um comportamento dinâmico não-linear à rede (HAYKIN, 2001). Estas redes também são conhecidas como redes dinâmicas recorrentes (HU; HWANG, 2002).

Neste contexto, uma vez especificada a quantidade de camadas e o número de neurônios em cada uma delas, deve ser definido o processo pelo qual ocorrerá o ajuste dos pesos sinápticos, definido pelo algoritmo de aprendizagem (BRAGA; LUDEMIR; CARVALHO, 2014; WILAMOWSKI; IRWIN, 2011).

### 3.3 ALGORITMOS DE APRENDIZAGEM

Uma das principais características das RNA é a sua habilidade de aprender através de exemplos. O treinamento é a forma com o qual a rede irá adaptar os seus pesos, objetivando reproduzir, da melhor maneira possível, o conhecimento adquirido pelos nós de entrada através de um conjunto de dados (BRAGA; LUDEMIR; CARVALHO, 2014). Haykin (2001) define a aprendizagem como um processo pelo qual os pesos sinápticos (ou parâmetros livres) são modificados por meio de um procedimento de estimulação a partir do ambiente no qual a rede está inserida. A maneira pela qual a modificação dos parâmetros ocorre, determina o tipo de aprendizagem realizada.

O conceito de aprendizagem está relacionado com a melhoria do desempenho da RNA seguindo algum critério de desempenho preestabelecido (BRAGA; LUDEMIR; CARVALHO, 2014). Na abordagem computacional, a aprendizagem é implementada a partir de algoritmos de treinamento, baseados em técnicas de otimização (BEALE; HAGAN; DEMUTH, 2017). Inicialmente, os pesos sinápticos são configurados aleatoriamente e conforme novas informações vão sendo inseridas, seus valores vão sendo alterados. De modo geral, o vetor de pesos  $\mathbf{w}(t + 1)$  no instante  $t + 1$  pode ser escrito conforme a Equação (27):

$$\mathbf{w}(t + 1) = \mathbf{w}(t) + \Delta\mathbf{w}(t) \quad (27)$$

sendo  $\mathbf{w}(t)$  o vetor de pesos no instante  $t$  e  $\Delta\mathbf{w}(t)$  o ajuste de peso aplicado.

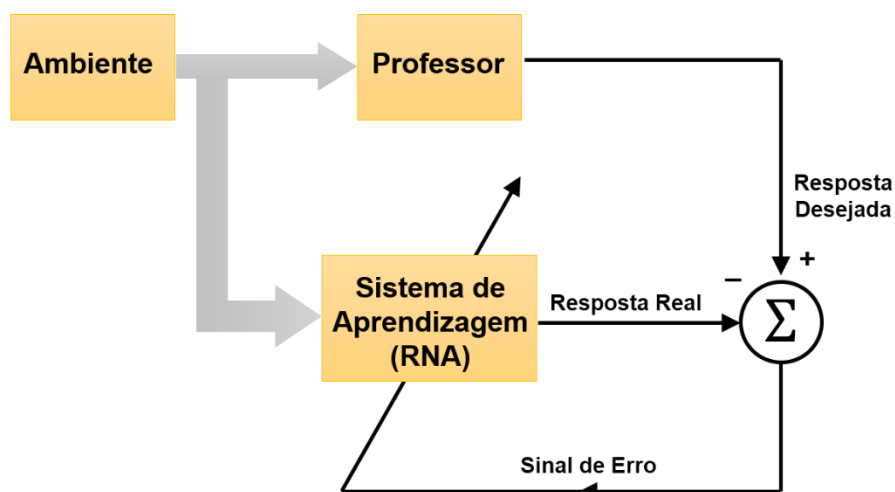
A maneira como se calcula  $\Delta\mathbf{w}(t)$  é o que determina o algoritmo de aprendizagem. Em geral, algoritmos de aprendizagem podem ser divididos em duas classes: aprendizado supervisionado e aprendizado não-supervisionado (BRAGA; LUDEMIR; CARVALHO, 2014). Em problemas de modelagem de sistemas e previsão, nos quais

a resposta desejada é conhecida, o algoritmo é considerado supervisionado (HAYKIN, 2001).

No treinamento supervisionado há a existência de um professor externo (supervisor), encarregado de alimentar as entradas da rede com um conjunto de padrões, observar a saída calculada pela mesma e compará-la com a saída desejada. Desta forma, como a resposta da rede está vinculada aos valores atuais do seu conjunto de pesos, estes são modificados fazendo com que a saída da rede se aproxime da saída desejada. Portanto, o supervisor é o responsável por fornecer as informações referentes ao direcionamento do ajuste dos pesos (BRAGA; LUDEMIR; CARVALHO, 2014).

Uma representação esquemática do aprendizado supervisionado é apresentada na Figura 17.

Figura 17 – Diagrama de blocos da aprendizagem supervisionada.



Fonte: O próprio autor.

De acordo com Braga, Ludemir e Carvalho (2014) e Haykin (2001), quaisquer algoritmos de aprendizagem supervisionado fundamentam-se na minimização do sinal de erro da resposta atual calculada pela rede em relação à saída desejada. Logo, sejam  $y(t)$  a resposta desejada e  $\hat{y}(t)$  a resposta atual computada pela rede, o sinal de erro pode ser calculado através da Equação (28):

$$e(t) = y(t) - \hat{y}(t) \quad (28)$$

O ajuste corretivo dos pesos é realizado visando reduzir a magnitude do sinal de erro  $e(t)$ . Este objetivo é então alcançado minimizando-se a função custo  $\varepsilon(t)$ , representada pelo Erro Quadrático Médio (*Mean Squared Error* – MSE), conforme a Equação (29):

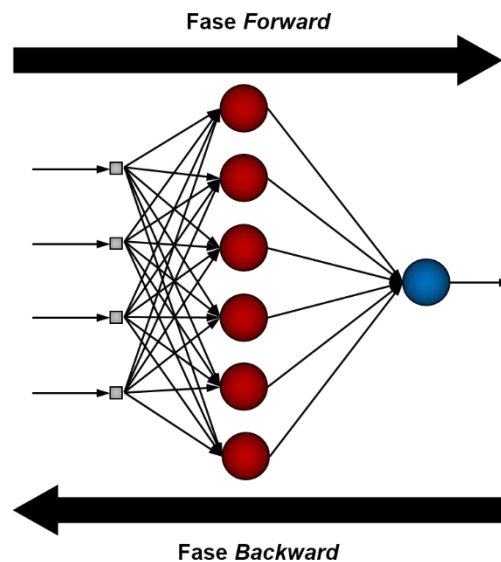
$$\varepsilon(t) = \frac{1}{n_T} \sum_{t=1}^{n_T} e^2(t) \quad (29)$$

em que  $n_T$  é o número de exemplos de treinamento e  $e^2(t)$  é o erro quadrático da saída. Este processo de aprendizagem é denominado aprendizagem por correção de erros e a minimização da função custo  $\varepsilon(t)$  resulta na regra de aprendizagem conhecida como Regra Delta ou Regra de Widrow-Hoff (WIDROW; HOFF, 1988).

Usando uma abordagem semelhante à regra de aprendizagem por correção de erro, o primeiro algoritmo para treinamento em múltiplas camadas foi o Algoritmo de Retropropagação de Erro ou *Back-Propagation* (RUMELHART; HINTON; WILLIAMS, 1986). Este algoritmo é bastante popular devido à sua simplicidade, embora o processo de treinamento seja muito lento (HERTZ; KROGH; PALMER, 1991; WILAMOWSKI; IRWIN, 2011).

O Algoritmo de *Back-Propagation* ocorre em duas fases, em que cada fase processa o sinal percorrendo as camadas da rede em um sentido distinto. A Figura 18 apresenta um esquemático do comportamento dessas duas fases, conhecidas como *forward* e *backward* (BRAGA; LUDEMIR; CARVALHO, 2014).

Figura 18 – Fluxo de processamento do Algoritmo de *Back-Propagation*.



Fonte: O próprio autor.

No passo para frente (fase *forward*), o sinal de entrada é propagado das entradas até a saída da rede e, a partir da resposta  $y(t)$  para um dado padrão de entrada  $x_{in}(t)$ , calcula-se o erro através da Equação (28). Por outro lado, durante a etapa de



retropropagação (fase *backward*), este sinal de erro é propagado para trás. Este sinal é o responsável pelo ajuste dos pesos entre as conexões da rede, fazendo com que os pesos sinápticos das camadas intermediárias sejam atualizados de forma que a saída da rede se aproxime estatisticamente da resposta desejada (BRAGA; LUDEMIR; CARVALHO, 2014; HAYKIN, 2001; HERTZ; KROGH; PALMER, 1991).

Contudo, conforme apresentado por Wilamowski e Irwin (2011) e Braga, Ludemir e Carvalho (2014), além de possuir um tempo de convergência lento, o desempenho deste algoritmo piora sensivelmente à medida que se aumenta a complexidade do problema a ser resolvido. Além disto, como este algoritmo emprega o método de gradiente decrescente, existe a possibilidade de ficar-se estacionado em um mínimo local. Visando solucionar tais inconvenientes, algumas variações deste algoritmo, como o algoritmo de Levenberg-Marquardt (LEVENBERG, 1944; MARQUARDT, 1963), foram desenvolvidas.

Diferentemente do algoritmo de treinamento *Back-Propagation*, que é baseado no método de gradiente decrescente, o algoritmo de Levenberg-Marquardt promove uma melhoria do método de Gauss-Newton através da variação da taxa de aprendizado (FUN; HAGAN, 1996). Em geral, todo algoritmo de treinamento evolui ao longo do tempo de forma similar, mas este algoritmo se diferencia dos demais por apresentar um desempenho muito bom logo no início de sua execução (BRAGA; LUDEMIR; CARVALHO, 2014; WILAMOWSKI; IRWIN, 2011).

### 3.4 REDES NEURAIIS COM ATRASO DE TEMPO

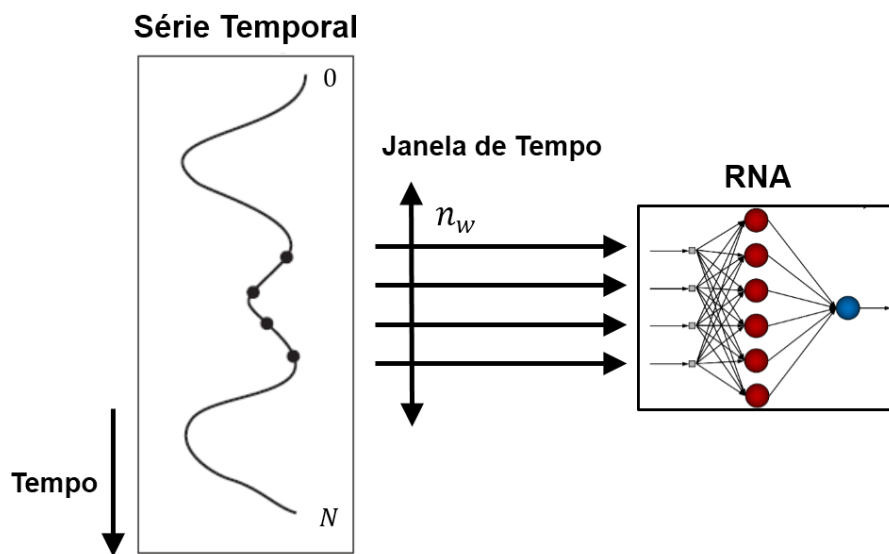
Uma RNA capaz de modelar a dinâmica de um processo, através de uma série temporal, é conhecida como rede neural dinâmica, do contrário, a rede é dita estática (PRINCIPE; EULIANO; LEFEBVRE, 2000). As arquiteturas de redes neurais dinâmicas são, em geral, extensões das Redes MLP, com a introdução de operadores de atraso unitário em cada uma das variáveis de entrada, objetivando-se, assim, representar a dinâmica (relações de causa e efeito) do processo não-linear que gerou a série temporal analisada (HU; HWANG, 2002).

A rede neural com atraso de tempo ou TDNN (*Time Delay Neural Networks*) é uma topologia de RNA amplamente empregada para a previsão de séries temporais pois utiliza os atrasadores de tempo para armazenar em *buffer* as variáveis de entrada atrasadas na camada de entrada (atrasadores externos) e/ou as saídas atrasadas na camada oculta (atrasadores internos), deslocando os dados de entrada e/ou saída

através de um janela de tempo de tamanho prefixado (HAYKIN, 2001; PRINCIPE; EULIANO; LEFEBVRE, 2000).

A Figura 19 ilustra como um número finito de atrasadores é capaz de percorrer toda a extensão da série temporal, a fim de converter um determinado sinal unidimensional  $\{x(t)\}_{t=0}^N$  em  $N - n_w - 1$  vetores de dimensão  $n_w + 1$ . Uma questão importante é como selecionar o comprimento da janela  $n_w$ , de forma que seja possível representar adequadamente as propriedades da série temporal.

Figura 19 – Exemplo de atrasadores formando uma janela de tempo de comprimento  $n_w$  na entrada de uma RNA.



Fonte: O próprio autor.

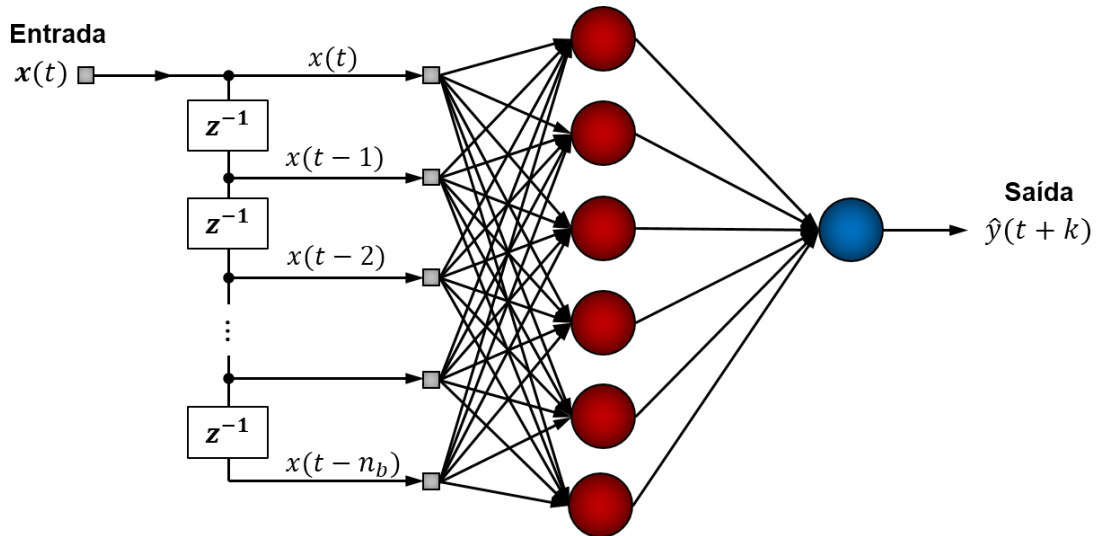
Deve-se ressaltar que a combinação de elementos de atraso unitário e pesos sinápticos associados faz com que as TDNN sejam funcionalmente equivalentes aos filtros de resposta ao impulso de duração finita (*Finite Impulse Response – FIR*), um dos blocos construtivos básicos no DSP (HAYKIN, 2001; HU; HWANG, 2002). Especificamente, conforme evidencia a Equação (30), esta topologia utiliza um conjunto de dados de entrada, contendo os  $n_b + 1$  valores observados de um sistema dinâmico ao longo do tempo,

$$\mathbf{x}(t) = [x(t), x(t-1), \dots, x(t-n_b)]^T \quad (30)$$

para prever uma determinada saída  $\hat{y}(t+k)$ , em que  $k$  é o horizonte de previsão.

A topologia de uma TDNN com uma camada oculta é apresentada na Figura 20. Nesta figura, o vetor de entrada é dado por uma sequência semelhante à Equação (30), com o intuito de mapear uma determinada saída  $k$  passos à frente (HAYKIN, 2001; PRINCIPE; EULIANO; LEFEBVRE, 2000). Esta topologia também é conhecida como *Nonlinear Input-Output* (NIO) (BEALE; HAGAN; DEMUTH, 2017).

Figura 20 – Arquitetura genérica de uma TDNN com uma única camada oculta.



Fonte: O próprio autor.

As TDNN também podem ser representadas através de um modelo não-linear autorregressivo com entradas exógenas (*Nonlinear AutoRegressive model with eXogenous inputs* – NARX), descrita pela Equação (31), com a inclusão de operadores de atraso também em sua saída. Assim, o modelo NARX por ser representado por:

$$\hat{y}(t) = \varphi[y(t-1), \dots, y(t-n_a); x(t), x(t-1), \dots, x(t-n_b)]^T \quad (31)$$

em que  $\hat{y}(t)$  é a saída estimada do modelo no instante  $t$ ,  $y(t)$  e  $x(t)$  representam, respectivamente, a entrada e saída do modelo,  $n_a$  é o número de atrasos da memória de saída,  $n_b$  é o número de atrasos da memória de entrada e  $\varphi(\cdot)$  representa uma função de ativação, em geral, não-linear, parametrizada através dos pesos sinápticos que são ajustados durante a fase de treinamento da rede (HAYKIN, 2001; LEONTARITIS; BILLINGS, 1985; NORGAAARD *et al.*, 2000).

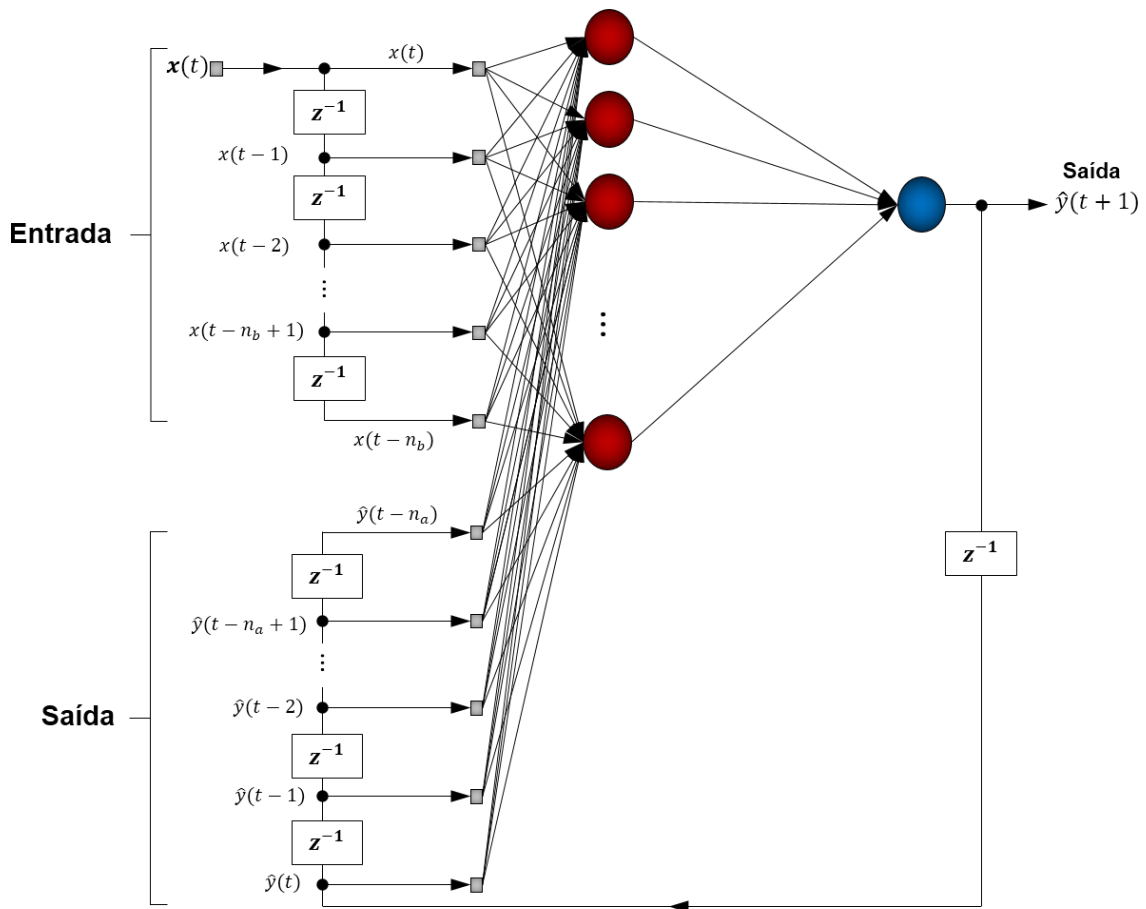
De acordo com Narendra e Parthasarathy (1990), a rede NARX pode ser configurada a partir de dois modos de operação: paralelo e série-paralelo. No modo de identificação paralelo (ver Figura 21), também chamado de modo recorrente, a saída

prevista é realimentada e incluída na saída do regressor, conforme a Equação (32). Por sua vez, de acordo com a Equação (33), no modo série-paralelo ou modo não-recorrente (ver Figura 22), a saída do regressor é composta somente pelos valores reais da saída do sistema.

$$\hat{y}(t) = \varphi[\hat{y}(t-1), \dots, \hat{y}(t-n_a); x(t), x(t-1), \dots, x(t-n_b)]^T \quad (32)$$

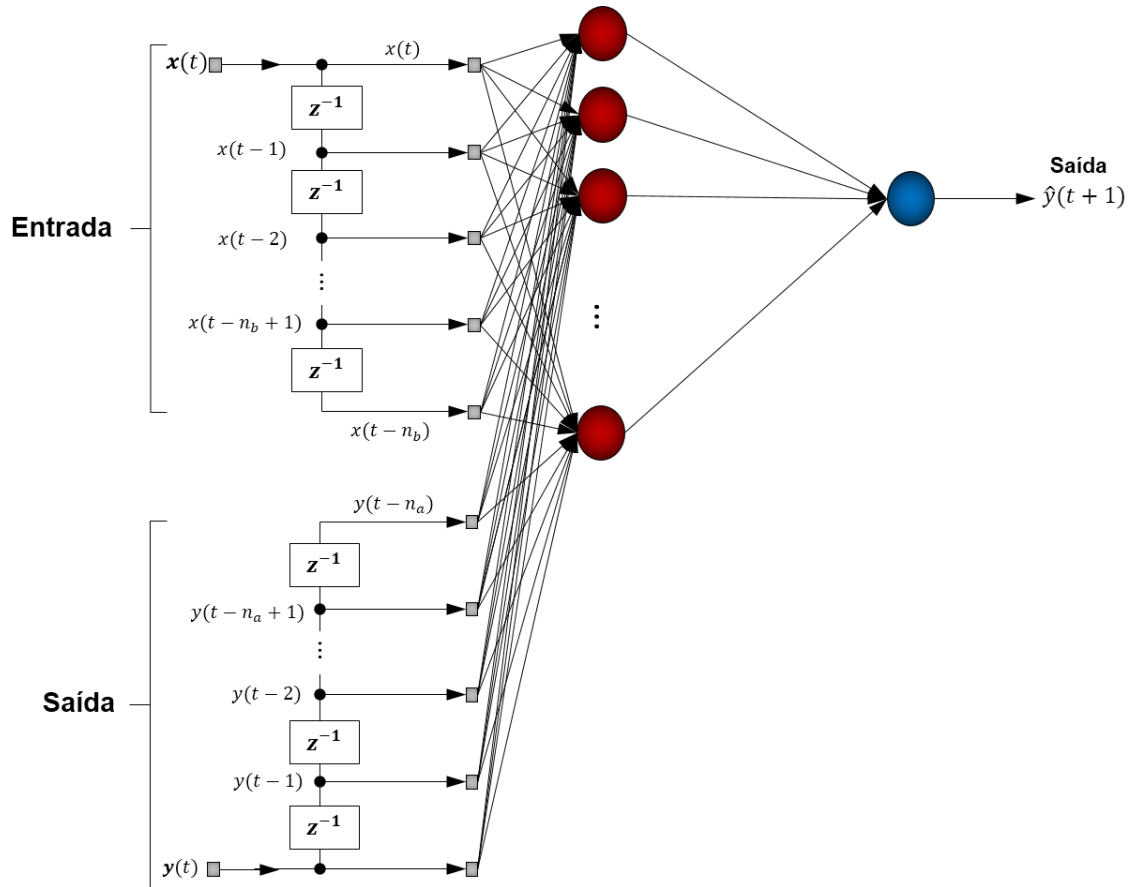
$$\hat{y}(t) = \varphi[y(t-1), \dots, y(t-n_a); x(t), x(t-1), \dots, x(t-n_b)]^T \quad (33)$$

Figura 21 – Rede NARX com uma única camada oculta, capaz de prever uma saída um passo à frente, configurada no modo de operação paralelo.



Fonte: O próprio autor.

Figura 22 – Rede NARX com uma única camada oculta, capaz de prever uma saída um passo à frente, configurada no modo de operação série-paralelo.



Fonte: O próprio autor.

À medida que avança o treinamento da rede no modo paralelo, as estimativas  $\hat{y}(t+k)$  tornam-se cada vez mais próximas dos valores desejados  $y_d(t+k)$ , indicando uma convergência do processo (BEALE; HAGAN; DEMUTH, 2017). No entanto, conforme apresentado por Narendra e Parthasarathy (1990), este modo de operação é muito mais instável do que o modo série-paralelo, devido, fundamentalmente, à incerteza propagada pelas estimativas realimentadas para o regressor de saída. Além disto, este tipo de operação também consome um elevado tempo de treinamento, em virtude das suas conexões realimentadas (HAYKIN, 2001; HU; HWANG, 2002).

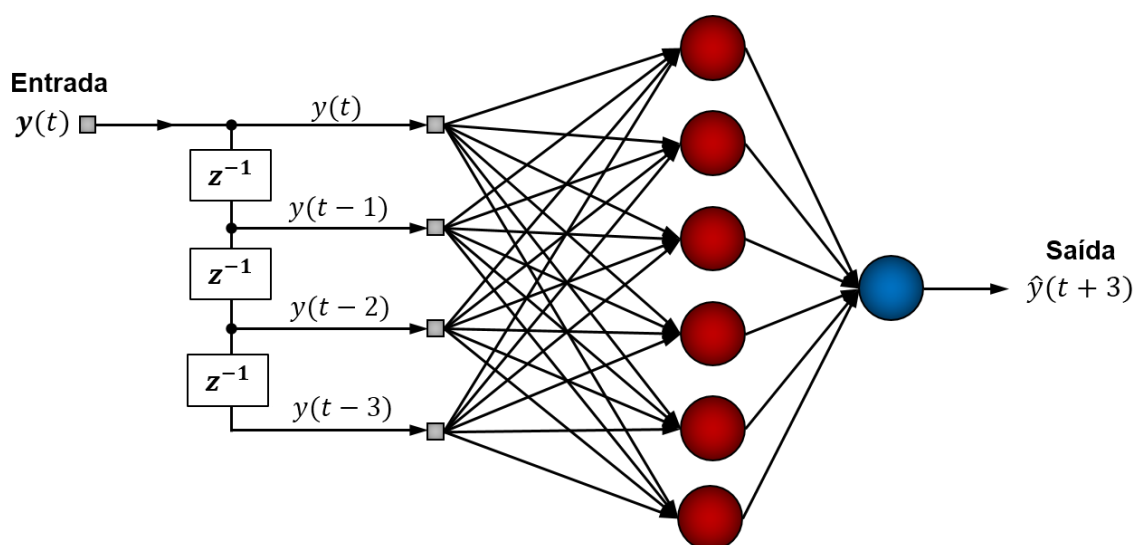
Outra forma de representar as TDNN é através do modelo não-linear autorregressivo (*Nonlinear AutoRegressive* – NAR). Esta rede pode ser entendida como uma versão simplificada da rede NARX, obtida pela eliminação da memória de entrada. Desta forma, esta rede realiza uma espécie de filtragem dinâmica, de forma que os valores passados da série temporal são usados para a previsão dos valores futuros da mesma (BEALE; HAGAN; DEMUTH, 2017).

Conforme é demonstrado pela Equação (34), de uma maneira geral, a saída estimada da NAR pode ser descrita através de seus  $n_a$  valores passados:

$$\hat{y}(t) = \varphi[y(t-1), \dots, y(t-n_a)] \quad (34)$$

A Figura 23 apresenta uma NAR capaz de prever uma determinada série temporal com um horizonte de previsão de 3 passos à frente, utilizando quatro valores passados da própria série (representados pelos 4 neurônios na camada de entrada) e uma única camada oculta contendo 6 neurônios.

Figura 23 – Exemplo de uma rede NAR com 6 neurônios na camada oculta e 4 neurônios de entrada para prever uma determinada saída 3 passos à frente.



Fonte: O próprio autor.

### 3.5 O PROJETO DE UMA RNA

Visto que as RNA podem ser aplicadas em diferentes tipos de tarefas (como por exemplo, aproximações de funções, previsão de séries temporais, classificação e reconhecimento de padrões), durante a sua concepção é necessário estabelecer qual o tipo de problema que está sendo trabalhado. Para cada tipo de tarefa, haverá uma arquitetura de rede e um tipo de aprendizado mais adequado (BRAGA; LUDEMIR; CARVALHO, 2014).

Depois de escolhida a arquitetura, a estrutura da rede é então determinada. Nesta etapa são definidos o número de camadas que formarão a rede, bem como o número de neurônios e o tipo de função de ativação que cada camada possuirá. O próximo passo é estabelecer de que forma os pesos serão inicializados e o algoritmo de treinamento a ser usado. Dependendo do conjunto de dados utilizado para o

treinamento, é necessário definir também a função custo responsável por medir o desempenho da rede (HAYKIN, 2001).

O projeto de uma rede neural também envolve um processo de validação do modelo obtido. Durante o treinamento iterativo da RNA, o conjunto de padrões é apresentado diversas vezes à rede e, a cada época, seus pesos são modificados. No entanto, encontrar o momento de parada ideal não é uma tarefa trivial, pois pode resultar em superdimensionamento (*overfitting*) ou em subdimensionamento (*underfitting*) da rede.

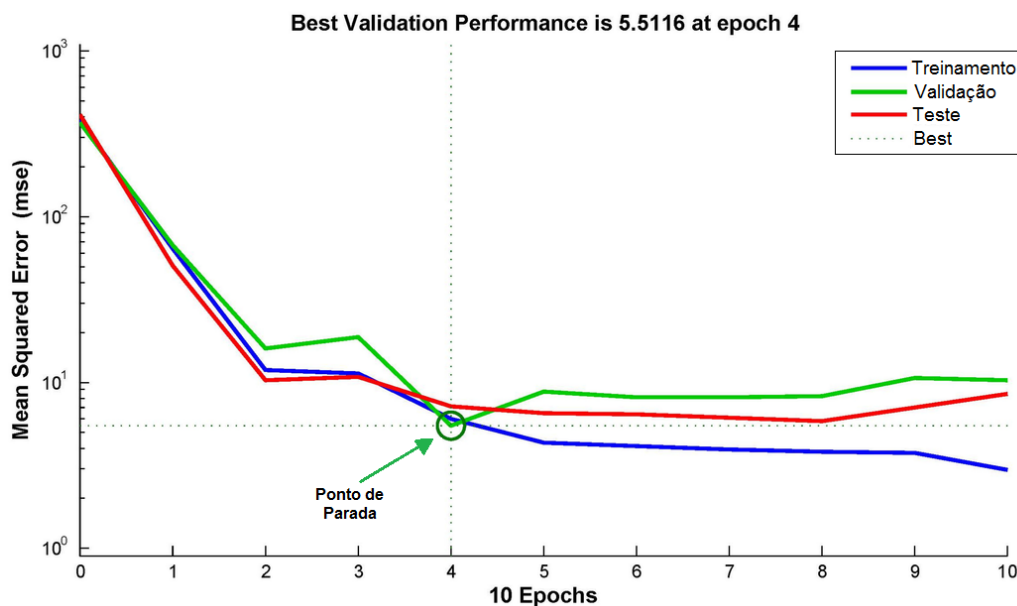
O *overfitting* ocorre quando a RNA decora os padrões, em virtude do excesso de treinamento, perdendo a sua capacidade de generalização. Caso a rede seja treinada menos que o necessário, ela não conseguirá aprender, ocasionando o *underfitting*. Sendo assim, uma forma de evitar tais impasses é a utilização da validação cruzada como critério de parada (BRAGA; LUDEMIR; CARVALHO, 2014).

De acordo com Haykin (2001), a validação cruzada divide o conjunto de dados em três parcelas: treinamento, validação e teste. A partir do conjunto de treinamento, os pesos da rede são ajustados seguindo o algoritmo de aprendizagem. Para avaliar a capacidade de generalização da rede durante o processo, utiliza-se o conjunto de validação. Como o ajuste dos pesos é feito a partir de um conjunto de dados distinto, o conjunto de validação sempre apresentará dados inéditos à rede. Diante disto, enquanto o erro do conjunto de validação estiver diminuindo, a rede continua generalizando.

O instante em que o erro de validação começa a aumentar, à medida que o erro de treinamento decrementa, é o indício de que a rede está perdendo a sua capacidade de generalização e o treinamento deve ser interrompido. Em seguida, utiliza-se o conjunto de testes para avaliar o desempenho da rede. É importante frisar que, caso durante o seu treinamento, a rede não consiga convergir para uma solução, ajustes poderão ser feitos tanto na estrutura, quanto nos parâmetros da mesma (HAYKIN, 2001).

A Figura 24 ilustra o melhor ponto de parada obtido através do algoritmo de validação cruzada. Nesta figura, são mostradas as três curvas de aprendizagem, relativas às medidas sobre os subconjuntos de treinamento, validação e teste.

Figura 24 – Critério de parada de validação cruzada: simulação.



Fonte: O próprio autor.

Uma outra maneira de se evitar o *overfitting* é a adoção de técnicas de poda (*prunning*), que, por sua vez, envolvem a eliminação de pesos e neurônios pouco significativos para a rede, facilitando assim, a resolução dos problemas de dimensionamento da rede e escolha de variáveis relevantes (BRAGA; LUDEMIR; CARVALHO, 2014; REED, 1993). Um detalhamento desta ferramenta é feito na Seção 3.6.

### 3.5.1 Medidas de Desempenho

As medidas de desempenho exercem um papel crítico na verificação da efetividade dos resultados obtidos pelo modelo. Tais medidas são utilizadas para verificar o nível de precisão que o modelo de previsão possui (BUENO, 2011). Segundo Lewis (1997), as medidas mais usadas para a verificação dos modelos de previsão são o MSE e o Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error – MAPE*).

O MSE é considerado uma importante medida para a escolha de um modelo de previsão mais preciso, uma vez que, os desvios pequenos resultam em erros pequenos, enquanto que os desvios mais relevantes sofrem maiores penalizações, conforme descreve a Equação (35) (LEWIS, 1997).

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 \quad (35)$$



em que  $N$  é o número total de amostras,  $y_t$  representa o valor real da amostra e  $\hat{y}_t$  refere-se ao valor estimado.

Já o MAPE é considerado uma das medidas de erro mais utilizadas na avaliação do modelo de previsão. De acordo com a Equação (36), o MAPE fornece uma indicação do tamanho médio do erro, expresso como uma percentagem do valor desejado  $y_t$ , que independe se o erro é positivo ou negativo (LEWIS, 1997).

$$MAPE (\%) = \frac{1}{N} \sum_{t=1}^N \frac{|y_t - \hat{y}_t|}{y_t} \cdot 100 \quad (36)$$

Observa-se pela Equação (36) que a saída desejada  $y_t$  não pode assumir valores nulos, isto pode ser corrigido através da normalização dos dados. O Percentual de Erro Absoluto (PEA) também é uma importante medida de desempenho pois é capaz de refletir a direção da mudança do processo, que é a referência mais relevante para a tomada de ações corretas (CHEN; WANG; HAN, 2010; YAN-JIONG; LIANG-TU; MIN, 1986). A Equação (37) evidencia o cálculo do PEA, isto é, o percentual de vezes em que o erro absoluto é menor de 0,05 em relação ao total de amostras.

$$PEA (\%) = \frac{1}{N} \sum_{t=1}^N N_E \cdot 100 \quad (37)$$

onde,

$$N_E = \begin{cases} 1, & \text{se } |y_t - \hat{y}_t| < 0,05 \\ 0, & \text{caso contrário} \end{cases}$$

### 3.6 ALGORITMOS DE PODA

Para a solução de problemas com redes neurais, frequentemente são necessárias o uso de redes altamente estruturadas e de tamanhos bastante elevados. Uma regra geral para que seja obtida uma boa generalização é a minimização do tamanho da rede, sem que o desempenho seja comprometido (HAYKIN, 2001; REED, 1993).

Em sua pesquisa, Sietsma e Dow (1988) demonstraram os benefícios da redução do número de neurônios em redes neurais previamente treinadas. Além da redução do tamanho operacional, estas redes aumentaram a sua capacidade de generalização e reduziram o efeito do ruído. Eles também sugerem algumas heurísticas para identificação dos neurônios que não contribuem substancialmente na composição da solução do problema.

Eles consideram que, se o neurônio tiver uma saída constante, para qualquer padrão de treinamento submetido ao conjunto de dados de entrada, então este neurônio contribui apenas com o *bias* e, portanto, não está participando da solução. Por sua vez, se as saídas de dois neurônios forem altamente correlacionadas (por exemplo, idênticas ou opostas) em todos os padrões, então, um deles é redundante e pode ser removido (SIETSMA; DOW, 1988).

Em geral, os Algoritmos de Poda podem ser agrupados em dois grupos. No primeiro, os algoritmos estimam a sensibilidade do índice MSE da função de erro (Equação (29)) em relação à remoção de um peso sináptico e, então, remove aqueles que apresentam menor influência. No segundo grupo, um termo adicional, chamado de termo de penalização, é acrescentado à função custo, de maneira a representar a complexidade da rede. É importante destacar que, existe uma sobreposição entre os grupos, pois a função custo pode conter termos relativos à sensibilidade (HAYKIN, 2001).

### 3.6.1 Poda da Rede baseada na Matriz Hessiana

Uma estratégia para a poda de redes neurais dinâmicas, baseadas em modelos entrada-saída, é a aplicação de algoritmos fundamentados no cálculo da matriz hessiana (NORGAARD *et al.*, 1996; NORGAARD, 2000), isto é, utilizando a informação sobre as derivadas de segunda ordem da superfície do erro. O objetivo desta técnica é estabelecer uma relação entre a complexidade da rede e o desempenho do erro de treinamento, a partir do cálculo da sensibilidade (HAYKIN, 2001).

Através de uma rede completamente conectada e com um determinado número de parâmetros, o suficiente para mapear as relações entrada-saída, é construído um modelo local da superfície do erro para prever analiticamente o efeito da retirada de cada peso sináptico sobre o erro da saída da rede. A partir de então, os pesos que apresentarem as menores influências sobre a rede são podados. O resultado é uma rede com menos pesos e parcialmente conectada (HAYKIN, 2001; LECUN *et al.*, 1990; REED, 1993).

LeCun *et al.* (1990) propõem a expansão por Série de Taylor para aproximação local da função custo, descrita de acordo com a Equação (38):

$$\varepsilon(\mathbf{w} + \Delta\mathbf{w}) = \varepsilon(\mathbf{w}) + \mathbf{g}^T(\mathbf{w})\Delta\mathbf{w} + \frac{1}{2}\Delta\mathbf{w}^T \mathbf{H}(\mathbf{w})\Delta\mathbf{w} + O(\|\Delta\mathbf{w}\|^3) \quad (38)$$

em que  $\Delta\mathbf{w}$  é a perturbação no vetor de pesos  $\mathbf{w}$  devido à retirada de um peso específico e  $\mathbf{g}(\mathbf{w})$  e  $\mathbf{H}(\mathbf{w})$  são, respectivamente, o vetor de gradiente e a matriz Hessiana, ambos calculados para o ponto de operação  $\mathbf{w}$ .

Assumindo que, os pesos sinápticos só serão eliminados após o processo de treinamento haver convergido, levando a rede a um mínimo global ou local da superfície do erro, o vetor de gradiente pode ser igualado a zero. Além disto, considerando que a superfície do erro em torno deste mínimo é aproximadamente quadrática, os termos de ordem mais elevadas também podem ser desconsiderados. Com estas duas suposições, a Equação (39) pode ser simplificada, de forma que, a variação do erro em função da remoção de um peso seja, então, dada pela Equação (48):

$$\Delta\varepsilon = \varepsilon(\mathbf{w} + \Delta\mathbf{w}) - \varepsilon(\mathbf{w}) \simeq \frac{1}{2}\Delta\mathbf{w}^T \mathbf{H}(\mathbf{w})\Delta\mathbf{w} \quad (39)$$

O objetivo é minimizar o aumento incremental de  $\varepsilon$  dado na Equação (39) sujeito a condição  $\Delta w_i + w_i = 0$ , isto é, sujeito a remoção do peso sináptico  $w_i$ . Para resolver este problema de otimização com restrições, utiliza-se uma solução a partir da construção do Lagrangiano, conforme a Equação (40):

$$S = \frac{1}{2}\Delta\mathbf{w}^T \mathbf{H}(\mathbf{w})\Delta\mathbf{w} - \lambda(\mathbf{1}_i^T \Delta\mathbf{w} + w_i) \quad (40)$$

onde  $\lambda$  é o multiplicador de Lagrange e  $\mathbf{1}_i^T$  é um vetor unitário cujos elementos são todos zeros, exceto o  $i$ -ésimo elemento, que é igual à unidade. Através da solução deste problema é possível obter uma expressão capaz de refletir a variação do erro devido a remoção do peso  $w_i$ , bem como a modificação ótima do vetor de peso  $\Delta\mathbf{w}$ , apresentadas, respectivamente, pelas Equações (41) e (42):

$$S_i = \Delta\varepsilon(i) = \frac{1}{2} \frac{w_i^2}{[\mathbf{H}(\mathbf{w})^{-1}]_{i,i}} \quad (41)$$

$$\Delta\mathbf{w} = \frac{w_i^2}{[\mathbf{H}(\mathbf{w})^{-1}]_{i,i}} \mathbf{H}(\mathbf{w})^{-1} \mathbf{1}_i \quad (42)$$

em que  $\mathbf{H}(\mathbf{w})^{-1}$  é a inversa da matriz Hessiana  $\mathbf{H}(\mathbf{w})$  e  $[\mathbf{H}(\mathbf{w})^{-1}]_{i,i}$  é o elemento  $(i, i)$  desta matriz. O Lagrangiano  $S_i$ , otimizado em relação a  $\Delta\mathbf{w}$ , sujeito à restrição que o

$i$ -ésimo peso sináptico  $w_i$  seja eliminado, é conhecido como a saliência (ou sensibilidade) do peso  $w_i$  (HAYKIN, 2001).

No método intitulado Dano Cerebral Ótimo (*Optimal Brain Damage* – OBD), LeCun *et al.* (1990) simplificam o cálculo das sensibilidades fazendo a suposição de que a Hessiana seja uma matriz diagonal, reduzindo substancialmente o esforço computacional. Entretanto, Hassibi, Stork e Wolff (1993) demonstram que computação completa da matriz Hessiana torna o método mais poderoso. Este procedimento é intitulado Cirurgião Cerebral Ótimo (*Optimal Brain Surgeon* – OBS).

Um resumo do processo de aplicação do OBS é apresentado como segue:

- Passo 1: Treinamento de uma arquitetura de rede totalmente conectada e com número de parâmetros o suficiente para capturar a relação entrada-saída;
- Passo 2: Ordenar os parâmetros em relação às suas respectivas sensibilidades;
- Passo 3: Eliminação do peso com menor sensibilidade e formação de uma nova configuração na arquitetura;
- Passo 4: Embora a Equação (42) seja empregada para a atualização do vetor de pesos da nova rede, é recomendável que a rede seja retreinada sempre que um peso ou uma parte deles (3 a 5%) for eliminado;
- Passo 5: Validação da rede;
- Passo 6: Enquanto a poda produzir resultados aceitáveis segundo algum critério previamente estabelecido, os procedimentos do Passo 2 em diante são repetidos.

Embora o cálculo da medida de sensibilidade, através do OBS, requeira maior esforço computacional, Hassibi, Stork e Wolff (1993) demonstram que este método é bastante poderoso, uma vez que não são feitas simplificações na Matriz Hessiana.

Em termos práticos, a tarefa de seleção de arquitetura de rede geralmente é realizada variando-se o número de nós e camadas no modelo. Desta forma, os algoritmos de poda podem ser aplicados para a seleção de uma arquitetura de rede capaz de evidenciar a relevância de cada nó, removendo aqueles identificados como redundantes ou não significativos, reduzindo a complexidade do modelo e melhorando a sua capacidade de generalização (CORANI; GUARISO, 2005).

#### 4 ESTUDO DE CASO

O teor de silício no ferro-gusa é um importante indicador do estado térmico do alto-forno, sendo capaz de refletir a qualidade do aço que está sendo produzido, bem como o quão eficiente está sendo a utilização da energia no processo de redução do ferro. Contudo, os operadores na usina em estudo só têm acesso aos percentuais de silício presente no ferro-gusa através de uma análise laboratorial atualizada a cada três horas, aproximadamente. Além disto, as condições hostis do processo não permitem que a avaliação do estado interno do alto-forno seja feita através de medições diretas. O alto-forno abordado por este estudo é considerado um equipamento de alta produtividade possuindo 4 furos de gusa, volume interno de  $3617\text{m}^3$ , 34 ventaneiras, cadinho com 12,5 metros de diâmetro, permitindo uma capacidade anual de produção de 2,8 milhões de toneladas de gusa (ARCELORMITTAL, 2017). A Figura 25 apresenta fotos desta planta.

Figura 25 – Alto-Forno 3 da ArcelorMittal Tubarão: a) Vista aérea; b) Torre do alto-forno.



Fonte: ArcelorMittal, 2017.

Em virtude da complexidade dos mecanismos e incerteza nas medidas do processo, os modelos fenomenológicos não apresentaram desempenho satisfatório na predição do comportamento dinâmico do teor de silício. Um grande problema é que os distúrbios internos, como mudanças temporárias nas taxas de fluxo locais de materiais durante o processo produtivo, exercem forte influência na troca de calor e na transferência de massa (SAXEN, 1994; FIGUEIRA; CASTRO; TAVARES, 1989).

Ao longo dos anos, foram publicados diversos trabalhos abordando o uso de técnicas lineares e não-lineares para a identificação dos mais variados modelos, em geral utilizando entradas exógenas, com o propósito de prever, e até mesmo controlar, o

teor de silício do ferro-gusa e demais parâmetros relativos à qualidade do processo de alto-forno (SAXEN; GAO, C.; GAO, Z., 2013).

#### 4.1 REVISÃO BIBLIOGRÁFICA

Os primeiros a aplicar sistematicamente os métodos de análise linear de séries temporais para a previsão de algumas das características do ferro-gusa (teor de silício, teor de enxofre e temperatura) foram Pandit, Clum e Wu (1975). Um modelo autorregressivo de ordem 1 foi desenvolvido utilizando 202 amostras de dados observados sobre o teor de enxofre, teor de silício, temperatura do ferro-gusa, taxa de ar de sopro e relação coque/minério. Analisando os resultados, os autores verificaram que as variáveis apresentavam interdependências consideráveis. A partir de então, os autores propuseram um esquema de previsão e controle das variáveis características do ferro-gusa.

No trabalho de Hashemi, Clum e Wu (1977) foi desenvolvida uma metodologia denominada *Dynamic Data System* (DDS), que descreve a conexão entre os modelos autorregressivos ao sistema correspondente. Foram utilizadas sete variáveis do alto-forno (volume, temperatura e umidade do forno, teor de silício, teor de enxofre e temperatura do ferro-gusa e a taxa de injeção de óleo). Após extensa análise dos resultados, os modelos foram avaliados através das perspectivas de previsão e controle. Uma previsão de um passo à frente do teor de silício mostrou uma precisão razoável do modelo, embora não tenha conseguido prever os picos mais acentuados do sinal.

Chung *et al.* (1986) também utilizaram a metodologia DDS para desenvolver um modelo de predição do teor de silício, empregando um modelo autorregressivo com 6 atrasos de tempo de cada uma das 10 variáveis de entrada: volume, temperatura, pressão de sopro, taxa de injeção de óleo, relação minério/coque, temperatura de chama, taxa de descida de carga, temperatura média do coque, permeabilidade do leito e observações anteriores do teor de silício. Como os resultados, a princípio, não foram satisfatórios, os autores adotaram um procedimento iterativo em que o atraso de tempo ótimo para cada variável de entrada era detectado minimizando uma função objetivo. Este procedimento resultou em um modelo com apenas três variáveis independentes, além da parte autorregressiva, capaz de prever um pouco melhor. As variáveis de entrada utilizadas pelo modelo final foram a temperatura média do coque, a taxa de injeção de óleo e a relação de minério/coque.

Yan-Jiong, Liang-Tu e Min (1986) apresentaram um estudo baseado na modelagem autorregressiva do teor de silício, destacando o fato de que não é apenas a precisão absoluta das previsões que deve ser levada em consideração como critério de desempenho, mas também, quantas vezes as previsões são precisas o suficiente (dentro de um erro permitido) e quantas vezes a direção do sinal previsto está de acordo com a do sinal verdadeiro. Para a análise, utilizaram 500 observações de dados de um alto-forno pequeno, com um intervalo de amostragem de cerca de 50 minutos. As variáveis de entrada foram a relação minério/coque, a taxa de injeção de carvão, o volume e a temperatura de sopro e uma temperatura teórica do coque nas ventaneiras. Usando o critério de Erro de Predição Final (*Final Prediction Error* - FPE), foi encontrada uma ordem ótima dos termos no modelo igual a três. Um modelo com valores de critérios de desempenho aceitáveis foi ilustrado, embora a previsão de um passo à frente tenha ficado atrasada, marginalmente melhor que uma previsão ingênua<sup>1</sup>.

Em geral, a irregularidade entre os intervalos de abertura e fechamento do furo de gusa é negligenciada pelos autores. Neste sentido, Saxén (1994) testou um modelo autorregressivo considerando a amostragem não uniforme das coletas de ferro-gusa, utilizando uma série de premissas simplificadoras do processo desenvolvidas por outros autores, de maneira que foi possível espaçar as informações disponível de teor de silício uniformemente ao longo do intervalo de tempo.

Com o intuito de filtrar os ruídos de alta frequência, foram usados filtros de média móvel com janela de 6 amostras às variáveis de entradas com medição uniforme (como parâmetros do ar de sopro e composição dos gases de topo), escolhidas por correlação cruzada com o teor de silício. Um diferencial deste trabalho foi a utilização da duração de cada corrida como entrada exógena do modelo. Em virtude das tendências não-estacionárias apresentadas pela evolução temporal do teor de silício, também foram testados modelos autorregressivos integrados. Esta diferenciação causou resultados mais ruidosos, embora tenha apresentado uma melhoria no desempenho das previsões durante alguns períodos (SAXEN, 1994).

Em geral, os modelos propostos foram capazes de seguir a tendência da série razoavelmente bem, para previsões variando de 1 à 25 passos à frente. Contudo, não

---

<sup>1</sup> De acordo com Barros (2004), a previsão ingênua é o método mais simples de previsão. Este método assume que a amostra a ser prevista referente ao tempo  $t + 1$ , será igual a última observação no tempo  $t$ .

houve uma análise mais criteriosa para a seleção das variáveis de entrada e muitas simplificações foram feitas com relação a dinâmica do alto-forno (SAXEN, 1994).

Partindo da premissa de que, em geral, os modelos contendo termos autorregressivos tende a tornar as previsões mais inertes (aproximando-se de um modelo ingênuo), Waller e Saxén (2000) estenderam o trabalho a uma análise mais rigorosa dos dados ao aplicar um modelo FIR, capaz de prever mudanças mais bruscas do teor de silício. Desta forma, baseando-se unicamente nas entradas do sistema, foi proposto um modelo adaptativo o qual pudesse estimar recursivamente seus parâmetros por meio de um Filtro de Kalman.

A princípio, utilizou-se o conhecimento da dinâmica do processo para a seleção de cerca de 40 potenciais variáveis de entrada, calculadas pela média das medições a cada 15 minutos. A partir de então, uma pesquisa exaustiva através de modelos de entrada simples com intervalos de tempo variando-se de 0 a 15 horas, detectou-se as entradas mais relevantes: um índice que reflete a reserva de energia no forno, uma variável que expressa o calor na zona de combustão, a taxa de produção e um índice de permeabilidade do leito (WALLER; SAXÉN, 2000).

Entretanto, como as variáveis são, em geral, índices com cálculo baseado nas mesmas medições, um certo grau de correlação foi inevitável e, por isso, diferentes possibilidades para combinar as entradas foram examinadas, sendo sugerido, ao final, um modelo com três entradas distintas e atrasos de 15 a 60 minutos. Ao comparar este modelo com o puramente autorregressivo, observou-se que o modelo FIR apresentou propriedades preditivas superiores, sendo capaz de não só prever as mudanças súbitas do sinal, como também estimar a duração dos distúrbios e paradas de alto-forno consideravelmente melhor (WALLER; SAXÉN, 2000).

Desta análise, observa-se que os modelos baseados em regressão apresentam algumas limitações, pois tendem a simplificar as relações temporais entre as variáveis do processo. Esta é a razão pela qual os modelos de séries temporais, que incorporam algum conhecimento temporal do processo foram relatados como sendo mais bem sucedidos. Contudo, a evidente não-linearidade apresentada pela evolução temporal da série de silício, torna as ferramentas não-lineares mais promissoras para realizar a previsão do teor de silício (SAXEN; GAO, C.; GAO, Z., 2013). Com o intuito de comprovar esta afirmação, Diniz *et al.* (2017) realizaram um comparativo entre as previsões um passo à frente do teor de silício no ferro-gusa empregando modelos lineares, do tipo ARIMA, e não-lineares, através de RNA do tipo NAR.



Durante os testes, observou-se que apesar do sinal previsto pelo modelo ARIMA ter conseguido acompanhar, mesmo com um pequeno atraso, o comportamento da série real, houveram pequenas discrepâncias nas amplitudes dos sinais. Por outro lado, embora a rede NAR tenha previsto com maior acurácia, foram observadas, em ambos os modelos, pequenas diferenças entre os sinais nos momentos em que houveram variações do percentual de silício acima de 0,4% (DINIZ, 2017).

Tunçkaya e Köklükaya (2016) também realizaram uma comparação entre os modelos obtidos através de redes neurais, modelos ARIMA e de Regressão Linear Múltipla (*Multiple Linear Regression* – MLR), utilizados para prognosticar e acompanhar a temperatura de um alto-forno, a partir de seis parâmetros do processo (fluxo do sopro de ar frio, temperatura de explosão à frio, oxigênio adicional, injeção de vapor adicional, taxa de injeção de carvão pulverizado e pressão de explosão) que afetam diretamente a variável em questão. Embora os resultados computacionais tenham sido satisfatórios, o modelo utilizando a rede neural foi o que apresentou o melhor desempenho.

As primeiras implementações de redes neurais na previsão de parâmetros do alto-forno ocorreram no final do século XX. Por exemplo, Bulsari, Saxén, H. e Saxén, B. (1992) usaram as redes *feedforward* de camadas múltiplas para predizer o teor de silício do ferro-gusa. Além disto, diversos modelos de RNA foram testadas por Singh, Sridhar e Deo (1996) para prever o teor de silício do ferro-gusa usando como entradas a taxa de coque, a temperatura de explosão quente, a taxa de escória, a pressão de topo, a basicidade de escória e o logaritmo da energia cinética de explosão.

Embora a maioria dos esforços de modelagem relatados nesta pesquisa bibliográfica tenha encontrado algum sucesso, ainda observam-se problemas fundamentais: a determinação das variáveis de entrada relevantes para o modelo entre um extenso conjunto de medidas disponíveis. A princípio, o problema de seleção de variáveis de entrada foi tradicionalmente abordado usando o conhecimento do processo, mas a escolha também deve incluir os atrasos de tempo pertinentes das variáveis. Desta forma, as mais variadas técnicas para a seleção das entradas e dos seus respectivos atrasos vêm sendo associadas às RNA para a solução deste tipo de problema (SAXEN; GAO, C.; GAO, Z., 2013).

Em pesquisas desenvolvidas por Dhond, Gupta e Vadhavkar (2000) e Bag (2007), a temperatura e a porcentagem do teor de silício do ferro-gusa também foram preditas utilizando as redes neurais. Com o intuito de remover os *outliers* do conjunto de dados,

os valores maiores do que dois desvios padrão da média foram modificados para que se adequassem ao intervalo e, em alguns casos, um valor mínimo para a variável foi especificado. Também foi utilizada a técnica da interpolação linear entre as medições para aproximar valores no caso de dados ausentes. Ademais, como o silício é medido com menor frequência em comparação com as variáveis de entrada, foi necessária uma reamostragem deste sinal. Para o pré-processamento das variáveis do modelo, também foi utilizada a técnica de janela móvel com 12 pontos e, a fim evitar grandes discrepâncias entre os valores das medições, foi aplicada uma normalização dos dados entre o intervalo de 0 a 1.

O atraso de cada variável de entrada foi ajustado e calculado por fatores de correlação: se o valor de correlação entre uma entrada em particular e a saída for alto, então essa variável seria relevante para a previsão. O valor limite do fator de correlação foi decidido com base na experiência de especialistas em alto-forno e na análise dos dados. Assim, partindo-se de um conjunto de dados contendo 39 potenciais variáveis de entrada, foram selecionadas 12 entradas para determinar a temperatura do ferro-gusa (coque total, óxido de carbono, hidrogênio, vapor, fluxo de calor do grupo 1, fluxo de calor do grupo 2, injeção de coque, porcentagem do enriquecimento de oxigênio, relação minério/coque, tempo de carga, temperatura de sopro e a temperatura do ferro-gusa anteriormente medida). Já para a previsão do teor de silício foi necessária a inclusão de 9 entradas adicionais relacionadas às características do coque e do sinter não especificadas pelos autores (BAG, 2007; DHOND; GUPTA; VADHAVKAR, 2000;).

No trabalho de Dhond, Gupta e Vadhavkar (2000), foram testados horizontes de previsão de 2 e 4 horas. A rede que apresentou o menor MSE possuía 19 nós ocultos, sendo capaz de prever as saídas com um grau de precisão elevado. Já no estudo realizado por Bag (2007), o horizonte de previsão utilizado foi de 6 e 8 horas. Os resultados obtidos foram comparados com os valores medidos e foi observado que, apesar de terem apresentado bons rendimentos, as precisões de ambas as estimativas diminuía conforme aumentava-se o horizonte de previsão.

Chen (2001) apresentou uma estratégia para prever o teor de silício com base no raciocínio qualitativo e nas redes neurais. Com a ajuda do conhecimento dos operadores do alto-forno estudado, foram identificadas as 10 variáveis mais relevantes para o modelo (volume, temperatura e pressão de sopro, quantidade de carvão, índice de desempenho de ventilação, pressão e temperatura do gás de topo,

quantidade de matéria-prima, taxa de coque e quantidade adicional de coque). O número de neurônios da camada oculta da rede neural foi determinado a partir de uma fórmula empírica. Deve-se destacar a constante atualização dos dados: toda vez que uma nova amostra do conjunto de teste estivesse disponível, a amostra mais antiga do conjunto de treinamento era descartada, de forma a manter o tamanho dos conjuntos inalterado.

Para avaliar o desempenho da previsão foram utilizados três critérios de desempenho: o MSE, o MAPE e o PEA menor que 0,1. As previsões de três passos à frente mostram-se surpreendentemente precisas durante todo o intervalo de tempo, embora não tenha sido esclarecido como foi possível obter a precisão relatada para sinais tão ruidosos (CHEN, 2001).

Saxén e Pettersson (2007) aplicaram o Algoritmo de Poda (*Prunning Algorithm*), em redes neurais, com o intuito de detectar não apenas as entradas relevantes ao problema de predição do silício como também os atrasos de tempo correspondentes. A partir deste algoritmo, eles apresentaram um modelo parcimonioso, isto é, as conexões pouco relevantes foram removidas gradualmente até que a eficiência da rede começou a ser comprometida. Esta pesquisa foi baseada em um conjunto de dados, no qual as variáveis foram pré-processadas para possuir uma amostragem contendo seus valores médios horários.

Inicialmente, foram selecionadas 15 potenciais variáveis de entrada (volume total, pressão de sopro, permeabilidade do gás, injeção de carvão, teor de  $O_2$ , consumo de CO, gás  $CO+CO_2$  de topo, temperatura de chama, taxa de coque, taxa de carvão, energia das ventaneiras, volume de sopro, relação  $CO/CO_2$ , perda de calor nas ventaneiras e relação minério/coque), descritas como séries autorregressivas, incluindo defasagens de até 8 horas. Uma análise estatística dos resultados obtidos através de várias execuções do Algoritmo de Poda demonstrou que as entradas encontradas são correlacionadas com o teor de silício. Além disto, os atrasos de tempo detectados mostraram-se razoáveis considerando a dinâmica do processo (SAXEN; PETTERSSON, 2007).

Os autores propuseram também uma aprendizagem *on-line* dos pesos da camada superior para adaptar o modelo às mudanças nas relações entrada-saída, utilizando-se um Filtro de Kalman. Esta atualização fez com que o modelo apresentasse um desempenho superior, sendo capaz de seguir mudanças bruscas do sinal e realizar previsões mais precisas utilizando um horizonte de previsão de dois passos à frente.

Embora este experimento tenha demonstrado que um modelo parcimonioso pode ser desenvolvido com base nos resultados do Algoritmo de Poda, a seleção final do modelo e a construção do mesmo foram feitas manualmente com base nos resultados (SAXÉN; PETTERSSON, 2007).

Visando resolver este problema, Nurkkala, Pettersson e Saxén (2011) aprimoraram a técnica de modelagem desenvolvida por Saxén e Pettersson (2007). O novo algoritmo era constituído por três estágios: primeiro, a partir de redes *Feedforward* do tipo Perceptron de múltiplas camadas foi aplicado o Algoritmo de Poda para detectar os possíveis candidatos à modelo. Em seguida, as redes que continham as combinações entre as conexões das entradas e os nós ocultos mais promissoras foram utilizadas para formação de uma Fronteira de Pareto, objetivando-se minimizar a complexidade do sistema e o erro de estimação.

As conexões foram então combinadas de forma ótima através da Programação Linear Inteira Mista (PLIM) para formarem um conjunto final de modelos, a partir do qual o usuário poderia selecionar aquele que melhor atendesse às suas necessidades. Apesar do modelo apresentado neste estudo ter sido capaz de prever alterações no teor de silício com até 3 horas de antecedência, uma falha apresentada por este método foi o longo tempo de execução em um PC padrão. A fase de poda utilizou cerca de uma semana para ser concluída e a fase de combinação durou cerca de dez dias (NURKKALA; PETTERSSON; SAXÉN, 2011).

Chen, Wang e Han (2010) desenvolveu um método híbrido o qual integra uma RNA com o Algoritmo Genético (AG) para prever o teor de metal quente de silício, por meio de 11 variáveis (taxa de carvão, taxa de coque, pressão de sopro, volume de sopro, temperatura de sopro, pressão de topo, percentual de carga, carregamento de enxofre, índice de qualidade de ventilação, basicidade da escória e o valor atual do teor de silício). A partir de então, aplicou-se o AG para a seleção dos pesos iniciais da RNA, embora que não tenham sido dados maiores detalhes se o AG foi utilizado para encontrar a melhor estrutura ou os parâmetros da rede. Para verificar a eficiência do sistema, foi feita uma comparação dos resultados previstos usando o modelo desenvolvido e o modelo convencional, sem a utilização do AG.

Os resultados finais não foram muito satisfatórios, pois o modelo possuía qualidade similar à uma previsão ingênua. Além disto, o tamanho da rede era relativamente grande: 11 entradas e 14 nós ocultos. Entretanto, o que chamou atenção neste estudo foi que depois da etapa de aprendizagem e predição, os autores buscaram controlar

o comportamento subsequente do silício com base no conhecimento adquirido pela rede, sendo muito útil para o monitoramento do sistema (CHEN; WANG; HAN, 2010). Portanto, em virtude da complexidade e do grande número de variáveis envolvidas no processo, é possível averiguar a grande dificuldade enfrentada pelos cientistas para a escolha de um algoritmo adequado que seja capaz de selecionar as entradas de maior relevância e, conseqüentemente, estabelecer relações entrada-saída de maneira eficaz. Além disto, sabe-se que os dados históricos do teor de silício do ferro-gusa, coletado do processo de alto-forno, são séries temporais não-lineares e não-estacionárias (DINIZ *et al.*, 2017; SAXEN, 1994), compostas por diferentes componentes de frequência (WANG *et al.*, 2015).

Tendo em vista todos estes fatores, Wang *et al.* (2015) adotaram um algoritmo para decompor a série silício em diferentes séries temporais de frequência, utilizando o *Empirical Mode Decomposition* (EMD). Tendo em vista a capacidade das RNA em aproximar relações não-lineares, cada uma das subséries fora prevista um passo à frente, através de uma rede neural dinâmica do tipo NAR. Por fim, a previsão do teor de silício da próxima amostra foi obtida a partir do somatório de cada uma das previsões. Os resultados deste experimento mostram que o método proposto também apresentou desempenho superior ao do método sem o uso da decomposição.

Na Tabela 1 são evidenciadas algumas das principais características dos trabalhos apresentados nesta seção:

Tabela 1 – Literaturas de Referência: Resumo (Continua).

	<b>Autores</b>	<b>Ano</b>	<b>Entradas Selecionadas</b>	<b>Técnica de Previsão</b>
1	Pandit, Clum e Wu	1975	Teor de enxofre, teor de silício, temperatura do ferro-gusa, taxa de ar de sopro e relação coque/minério.	Modelo linear autorregressivo com entradas exógenas de ordem 1.
2	Hashemi, Clum e Wu	1977	Volume, temperatura e umidade do forno, teor de silício, teor de enxofre e temperatura do ferro-gusa e a taxa de injeção de óleo.	Modelo linear autorregressivo com entradas exógenas + DDS.
3	Chung <i>et al.</i>	1986	Volume, temperatura, pressão de sopro, taxa de injeção de óleo, relação minério/coque, temperatura de chama, taxa de descida de carga, temperatura média do coque, permeabilidade do leito e observações anteriores do teor de silício.	Modelo linear autorregressivo com entradas exógenas de ordem 6 + DDS.

Tabela 1 – Literaturas de Referência: Resumo (Continuação).

	<b>Autores</b>	<b>Ano</b>	<b>Entradas Seleccionadas</b>	<b>Técnica de Previsão</b>
4	Yan-Jiong, Liang-Tu e Min	1986	Relação minério/coque, a taxa de injeção de carvão, o volume e a temperatura de sopro e uma temperatura teórica do coque nas ventaneiras.	Modelo linear autorregressivo com entradas exógenas.
5	Saxén	1994	Perdas de calor, pressão de sopro, índice de permeabilidade e duração da corrida.	Modelo linear autorregressivo com entradas exógenas.
6	Waller e Saxén	2000	Reserva de energia no forno, uma variável que expressa o calor na zona de combustão, a taxa de produção e um índice de permeabilidade do leito.	Modelo FIR + Filtro de Kalman.
7	Diniz <i>et al.</i>	2017	Observações anteriores do teor de silício	Comparativo entre Modelos lineares do tipo ARIMA e RNA do tipo NAR.
8	Tunçkaya e Köklükaya	2016	Fluxo do sopro de ar frio, temperatura de explosão à frio, oxigênio adicional, injeção de vapor adicional, taxa de injeção de carvão pulverizado e pressão de explosão.	Comparativo entre Modelos lineares do tipo ARIMA, MLR e RNA..
9	Singh, Sridhar e Deo	1996	Taxa de coque, a temperatura de explosão quente, a taxa de escória, a pressão de topo, a basicidade de escória e o logaritmo da energia cinética de explosão.	Modelos empregando RNA.
10	Dhond, Gupta e Vadhavkar	2000	Coque total, óxido de carbono, hidrogênio, vapor, fluxo de calor do grupo 1, fluxo de calor do grupo 2, injeção de coque, porcentagem do enriquecimento de oxigênio, relação minério/coque, tempo de carga, temperatura de sopro, temperatura do ferro-gusa anteriormente medida, mais 9 entradas adicionais relacionadas às características do coque e do síter não especificadas.	Modelos empregando RNA.
11	Bag	2007	Coque total, óxido de carbono, hidrogênio, vapor, fluxo de calor do grupo 1, fluxo de calor do grupo 2, injeção de coque, porcentagem do enriquecimento de oxigênio, relação minério/coque, tempo de carga, temperatura de sopro, temperatura do ferro-gusa anteriormente medida, mais 9 entradas adicionais relacionadas às características do coque e do síter não especificadas.	Modelos empregando RNA.

Tabela 1 – Literaturas de Referência: Resumo (Conclusão).

Autores		Ano	Entradas Seleccionadas	Técnica de Previsão
12	Chen	2001	Volume, temperatura e pressão de sopro, quantidade de carvão, índice de desempenho de ventilação, pressão e temperatura do gás de topo, quantidade de matéria-prima, taxa de coque e quantidade adicional de coque.	Modelos empregando RNA + raciocínio qualitativo.
13	Saxén e Pettersson	2007	Volume total, pressão de sopro, permeabilidade do gás, injeção de carvão, teor de O <sub>2</sub> , consumo de CO, gás CO+CO <sub>2</sub> de topo, temperatura de chama, taxa de coque, taxa de carvão, energia das ventaneiras, volume de sopro, relação CO/CO <sub>2</sub> , perda de calor nas ventaneiras e relação minério/coque.	Modelos empregando RNA + Algoritmo de Poda.
14	Nurkkala, Pettersson e Saxén	2011	Volume total, pressão de sopro, permeabilidade do gás, injeção de carvão, teor de O <sub>2</sub> , consumo de CO, gás CO+CO <sub>2</sub> de topo, temperatura de chama, taxa de coque, taxa de carvão, energia das ventaneiras, volume específico, perda de energia, perda de calor nas ventaneiras, parâmetro referente a matéria-prima e relação minério/coque.	Modelos empregando RNA + Algoritmo de Poda + PLIM.
15	Chen, Wang e Han	2010	Taxa de carvão, taxa de coque, pressão de sopro, volume de sopro, temperatura de sopro, pressão de topo, percentual de carga, carregamento de enxofre, índice de qualidade de ventilação, basicidade da escória e o valor atual do teor de silício.	Modelos empregando RNA + Algoritmo Genético.
16	Wang <i>et al.</i>	2015	Observações anteriores do teor de silício.	EMD + RNA do tipo NAR.

Fonte: O próprio autor.

Analisando os trabalhos desenvolvidos, ao longo dos anos, implementando modelos para prever o teor de silício no ferro-gusa, observa-se a predominância de algumas variáveis de processo utilizadas como entradas exógenas, tais como a taxa de coque, relação minério/coque, volume de ar de sopro, temperatura do ar de sopro, pressão do ar de sopro, pressão no topo e temperatura do ferro-gusa (ver Tabela 2). No entanto, em virtude do grande número de variáveis envolvidas no processo e que, potencialmente influenciam a variável de saída, a maioria dos trabalhos destaca a dificuldade na escolha das entradas exógenas, bem como dos seus respectivos atrasos, durante a concepção do modelo (BAG, 2007; CHEN, 2001; CHEN; WANG;

HAN, 2010; DHOND; GUPTA; VADHAVKAR, 2000; NURKKALA; PETTERSSON; SAXÉN, 2011; SAXÉN; PETTERSSON, 2007; SAXEN; GAO, C.; GAO, Z., 2013; SINGH; SRIDHAR; DEO, 1996).

Tabela 2 – Entradas empregadas nos modelos implementados nas literaturas de referência (Continua).

Variáveis	Autores															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Basicidade da escória									x						x	
Calor na zona de combustão						x										
Consumo de CO													x	x		
Coque total										x	x					
Duração da corrida					x											
Energia das ventaneiras													x	x		
Fluxo de calor do grupo 1										x	x					
Fluxo de calor do grupo 2										x	x					
Fluxo do sopro de ar frio								x								
Gás CO+CO <sub>2</sub> de topo													x	x		
Hidrogênio										x	x					
Índice de desempenho de ventilação												x			x	
Índice de permeabilidade					x	x							x	x		
Injeção de vapor adicional								x								
Logaritmo da energia cinética de explosão									x							
Óxido de carbono										x	x					
Oxigênio adicional								x								
Perda de calor nas ventaneiras																
Perdas de calor					x											
Permeabilidade do leito			x													
Porcentagem do enriquecimento de oxigênio										x	x		x	x		
Pressão de explosão								x								
Pressão de sopro					x							x	x	x	x	
Pressão de topo									x							
Pressão do gás de topo												x			x	
Quantidade adicional de coque												x				
Quantidade de carvão												x	x	x		
Quantidade de matéria-prima												x				
Relação CO/CO <sub>2</sub>													x	x		
Relação minério/coque	x		x	x						x	x		x	x	x	



Tabela 2 – Entradas empregadas nos modelos implementados nas literaturas de referência (Conclusão).

Variáveis	Autores															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Reserva de energia no forno						x										
Taxa de ar de sopro	x															
Taxa de coque									x	x	x	x	x	x	x	
Taxa de descida de carga			x													
Taxa de escória									x							
Taxa de injeção de carvão pulverizado				x				x					x	x	x	
Taxa de injeção de óleo		x	x													
Taxa de produção						x										
Temperatura de chama			x										x	x		
Temperatura de explosão à frio								x								
Temperatura de explosão quente									x							
Temperatura de sopro			x	x						x	x	x				x
Temperatura do ferro-gusa	x	x								x	x					
Temperatura do forno		x														
Temperatura do gás de topo												x				
Temperatura média do coque			x													
Temperatura teórica do coque nas ventaneiras				x												
Tempo de carga												x				
Teor de enxofre	x	x														x
Teor de silício	x	x	x				x								x	x
Umidade de sopro		x	x													
Umidade do forno		x														
Vapor de calor do grupo 1										x	x					
Volume de sopro			x	x								x	x	x	x	
Volume do forno		x											x	x		

Fonte: O próprio autor.

## 4.2 PRINCIPAIS VARIÁVEIS ENVOLVIDAS NO PROCESSO

A Tabela 3 apresenta as 28 potenciais variáveis de entrada, selecionadas a partir de uma criteriosa análise da literatura existente, bem como pela avaliação dos engenheiros especialistas e operadores do alto-forno estudado. Nesta tabela também são descritos os valores máximo e mínimo que cada uma destas variáveis podem assumir durante os períodos de estabilidade do processo. Valores fora desta faixa

podem ser considerados *outliers* e passarão por tratamento específico. Outrossim, são informados o tempo médio de resposta no processo de cada uma das variáveis.

Tabela 3 – Potenciais variáveis de entrada do modelo e suas principais características (Continua).

Nº	Variável de Entrada	Unidade de Medida	Valor Mínimo	Valor Máximo	Tempo Médio de Resposta
1	Temperatura de Chama	°C	2050	2250	2 a 3h
2	Relação CO/CO <sub>2</sub>	-	0,9	1,1	2 a 3h
3	Eficiência do gás H <sub>2</sub>	%	35	50	2 a 3h
4	Velocidade de Produção de Gusa	ton/min	4,5	6,5	2 a 3h
5	Taxa de Coque	kg/ton	260	340	6 a 8h
6	Taxa de injeção de carvão pulverizado ( <i>PCI Rate</i> )	kg/ton	140	220	2 a 3h
7	Redução Direta	%	27	33	2 a 3h
8	Relação Minério/Coque	-	4,5	5,5	6 a 8h
9	Índice Térmico H0	10 <sup>3</sup> kcal/ton	450	600	2 a 3h
10	IFGCN: Índice de Fluxo Gasoso Central Norte	-	3	6	2 a 3h
11	IFGPN: Índice de Fluxo Gasoso Periférico Norte	-	1	1,6	2 a 3h
12	IFGCS: Índice de Fluxo Gasoso Central Sul	-	3	6	2 a 3h
13	IFGPS: Índice de Fluxo Gasoso Periférico Sul	-	1	1,6	2 a 3h
14	Volume do ar soprado pelas ventaneiras (Volume de Sopro)	Nm <sup>3</sup> /min	4700	5100	2 a 3h
15	Umidade do ar soprado pelas ventaneiras (Umidade de Sopro)	g/Nm <sup>3</sup>	15	30	2 a 3h
16	Taxa (de Enriquecimento) de O <sub>2</sub>	%	4	7,5	2 a 3h
17	Temperatura do ar soprado pelas ventaneiras (Temperatura de Sopro)	°C	1190	1250	2 a 3h
18	Pressão do ar soprado pelas ventaneiras (Pressão de Sopro)	Kg/cm <sup>2</sup>	3,6	4,2	2 a 3h
19	Pressão (do gás) de Topo	Kg/cm <sup>2</sup>	2,4	2,5	2 a 3h
20	Temperatura (do gás) de Topo	°C	90	180	2 a 3h
21	Teor de H <sub>2</sub> no topo	%	4	5,5	2 a 3h

Tabela 3 – Potenciais variáveis de entrada do modelo e suas principais características (Conclusão).

Nº	Variável de Entrada	Unidade de Medida	Valor Mínimo	Valor Máximo	Tempo Médio de Resposta
22	Teor de N <sub>2</sub> no topo	%	45	50	2 a 3h
23	Teor de CO no topo	%	23	26	2 a 3h
24	Teor de CO <sub>2</sub> no topo	%	23	26	2 a 3h
25	Eficiência do gás CO	%	48	52	2 a 3h
26	Perdas Térmicas pelo <i>Stave</i>	10 <sup>4</sup> Kcal/h	1200	2500	2 a 3h
27	Temperatura do ferro-gusa	°C	1480	1530	2 a 3h
28	Força-Eletromotriz	%	35	55	2 a 3h

Fonte: O próprio autor.

As variáveis IFGCN (ou IFGCS) e IFGPN (ou IFGPS) são índices calculados a partir das temperaturas captadas através de sondas inseridas acima da carga, que refletem, respectivamente, as marchas centrais e periféricas do forno. Assim, um aumento da marcha indica uma maior circulação da carga em estado líquido em direção ao fundo do cadinho (FIGUEIRA, 1989).

De acordo com a avaliação dos especialistas e pelo estudo descrito por Figueira (1989), é possível fazer algumas afirmações acerca das variáveis IFGCN, IFGCS, IFGPN e IFGPS. Estes índices foram escolhidos, uma vez que, as variações na geometria e localização da zona coesiva podem influenciar no teor de silício produzido. Quanto maior for o IFGCN, mais alongada será o centro da zona coesiva, enquanto que, o valor do IFGCS indica o quanto este centro está deslocado. Em relação aos índices IFGPN e IFGPS, quanto maior forem as suas magnitudes, maior será a área de contato da zona coesiva, respectivamente, com as paredes norte e sul do forno.

Além dos parâmetros de temperatura, também são utilizados outros parâmetros como a Redução Direta, Perdas Térmicas pelo *Stave* e o Índice Térmico H0 para indicar o estado térmico do alto-forno, sendo as duas primeiras variáveis inversamente proporcionais e a última diretamente proporcional ao aporte térmico.

A Redução Direta representa o consumo de energia necessário para que o óxido de ferro seja diretamente reduzido a ferro metálico através da reação com o carbono proveniente dos combustíveis sólidos (coque ou carvão). Já o Índice Térmico H0 refere-se ao balanço térmico no interior do forno, calculado a partir das reações de

combustão nas ventaneiras junto ao calor sensível proveniente do sopro e da perda térmica de calor ocorrida por meio das reações endotérmicas, como a redução direta, decomposição da água, reações com hidrogênio e perdas por calcinação.

A Velocidade de Produção de Gusa (VPG) reflete o nível de produtividade. Esta produtividade é afetada diretamente pelo volume de gás que atravessa a carga que, por sua vez, é diretamente proporcional ao consumo de coque (Taxa de Coque). Com o aumento da Taxa de Coque, menor o tempo gasto para o aquecimento e amolecimento da carga metálica (tempo de residência) e, conseqüentemente, mais elevada ficará a localização da zona coesiva. Figueira (1989) afirma que, mantendo-se todos os outros parâmetros constantes, quanto mais elevada for a posição da zona coesiva, maior o percentual de silício gerado.

A distribuição da carga também é um fator que influencia no teor de silício produzido. Esta distribuição pode ser avaliada através da Relação Minério/Coque no sentido radial do forno: com o aumento da Relação Minério/Coque, menor será a permeabilidade à passagem do gás, por conseguinte, maior será a espessura da zona coesiva. Com este aumento, maior será o tempo de residência do gusa em contato com o gás SiO e, portanto, mais silício será incorporado ao gusa.

Outros fatores que também afetam a posição da zona coesiva são a velocidade do ar injetado pelas ventaneiras e as injeções de combustíveis auxiliares. Uma maior vazão de ar, para uma mesma área de ventaneira, resulta em maiores volumes de ar, implicando em uma maior penetração do sopro.

Por outro lado, a injeção de combustível auxiliar, como o carvão pulverizado (*PCI Rate*), ocasiona um acréscimo do volume de gás escoando através da carga e uma diminuição da vazão mássica dos sólidos, devido à substituição parcial do coque carregado no topo. Isto fará com que a zona coesiva seja descolada em direção ao topo.

Já o aumento da injeção de oxigênio (aumento da Taxa de Enriquecimento de O<sub>2</sub>), ocasiona uma redução no volume dos gases, devido à redução no volume de nitrogênio (N<sub>2</sub>) do ar soprado, o que provoca um deslocamento da zona coesiva em direção ao fundo do forno, ocasionando em uma diminuição da incorporação de silício ao gusa. No entanto, é importante lembrar também que, com o aumento da umidade do ar de sopro, as injeções de O<sub>2</sub> podem elevar a Temperatura de Chama e, conseqüentemente, gerar maior quantidade de SiO.

Um aumento na Pressão de Sopro diminui o percentual de silício incorporado ao gusa, pois promove uma menor formação de  $\text{SiO}$ , como também inibe a reação de redução direta do  $\text{FeO}$  e  $\text{MnO}$ , aumentando-se assim, os níveis de oxidação do silício ao atravessar a escória. As Temperaturas de Sopro e Temperatura de Topo também são importantes pois, em conjunto, são capazes de refletir o fluxo térmico do forno. Além disto, esta última variável é afetada diretamente pela umidade do coque. Já a pressão de topo é utilizada para o monitoramento da velocidade do gás na cuba e do tempo de residência do gás na carga.

No entanto, apesar de apresentarem evidentes correlações com o percentual de silício no ferro-gusa, deve ser frisado que, as variáveis pré-selecionadas não devem ser analisadas separadamente, uma vez que, devido à complexidade da dinâmica do alto-forno, a ação de uma variável em conjunto com as demais, pode influenciar positivamente ou negativamente a ação de outra.

Alguns conjuntos de variáveis e índices, como a Teor de  $\text{H}_2$  no topo e a Eficiência do gás  $\text{H}_2$ , bem como a Teor de  $\text{N}_2$  no topo, a Teor de  $\text{CO}$  no topo, a Teor de  $\text{CO}_2$  no topo, a Eficiência do gás  $\text{CO}$  e a Relação  $\text{CO}/\text{CO}_2$ , que mensuram os níveis de composição química dos gases e/ou o quanto foram consumidos, são, de alguma forma, proporcionais entre si. As redundâncias diretas ou indiretas entre as variáveis são necessárias, uma vez que também em virtude da complexa dinâmica do processo, não se sabe *a priori* qual ou quais parâmetros serão mais representativos para o modelo. Neste contexto, o Algoritmo de Poda pode ser utilizado para a solução deste tipo de problema, posto que o mesmo pode atuar na eliminação de tais redundâncias, ficando resultantes apenas os parâmetros que melhor representem a dinâmica a ser modelada.

De acordo com os especialistas e operadores do alto-forno estudado, em geral, o tempo de residência da carga no interior do forno é de cerca de 6 horas, a amostra atual do teor de silício pode estar correlacionada com 6 amostras passadas, considerando-se um tempo de amostragem de 60 minutos.

Com relação às informações pertinentes as corridas de ferro-gusa, que acompanham a série de silício, nota-se que a cada corrida ocorrem de dois a três carregamentos de carro torpedo e que, normalmente, em cada carregamento uma amostra do ferro-gusa é coletada no canal de gusa para análise laboratorial. Posteriormente, o percentual de silício encontrado nesta análise, referente ao determinado carregamento, é atualizado

no sistema. Todos estes fatores colaboram para a irregularidade do tempo de amostragem do percentual de silício, que varia em torno de 50 minutos.

Além disto, em virtude da logística de transporte das amostras, estas são encaminhadas ao laboratório a cada, no mínimo, 3 horas. Cada análise demora cerca de 3 a 5 minutos, no entanto, como várias amostras chegam ao laboratório simultaneamente, a análise é feita em batelada e os respectivos resultados são atualizados no sistema automaticamente. Portanto, devido à dinâmica atual do processo de coleta/análise das amostras de ferro-gusa, as atualizações no sistema de novas amostras de percentual de silício ocorrem a cada, no mínimo, 3 horas.

Outras variáveis como a Temperatura do Ferro-gusa e a Força-Eletromotriz<sup>2</sup> (*ElectroMotive Force* – EMF) também são mensuradas a cada carregamento de carro torpedo e, por isso, também apresentam mesmo período de amostragem irregular que o teor de silício. As demais variáveis descritas na Tabela 3 possuem uma resolução de 30 minutos.

Outro fator que deve ser destacado é o motivo da não utilização de dados como, por exemplo, a basicidade da escória e demais composições químicas do gusa, que apresentam forte correlação com o percentual de silício no ferro-gusa. Assim como o teor de silício, estes dados são obtidos a partir da análise laboratorial e, por conseguinte, também demandam de um longo tempo para a sua atualização no sistema, o que poderia comprometer a eficiência da previsão do modelo durante a sua utilização *on-line*.

---

<sup>2</sup> Sinal medido na carcaça do alto-forno, utilizado como indicativo do nível dos líquidos no cadinho e que reflete o estado térmico e químico no interior do forno (Gomes, 2016).

## 5 DESENVOLVIMENTO DE MODELOS E RESULTADOS

Este trabalho visa a elaboração de uma ferramenta capaz de prever o teor de silício no ferro-gusa através de dados históricos desta variável e/ou baseando-se em demais parâmetros que influenciam de forma significativa no valor presente do silício, fazendo uso de redes neurais com atraso de tempo.

Neste capítulo serão apresentados os experimentos, acompanhados das suas respectivas análises. Os procedimentos aqui evidenciados serão baseados a partir de variadas metodologias que, conforme descritos nas literaturas de referência, propiciaram a modelagem e previsão do teor de silício no ferro-gusa na saída do alto-forno. Também serão incluídas algumas adaptações visando melhor representar o processo do alto-forno em questão.

### 5.1 METODOLOGIA

Para o desenvolvimento deste estudo, serão utilizados dados oriundos do Alto-Forno 3 da ArcelorMittal Tubarão, localizada na região da Grande Vitória, Espírito Santo, referentes ao período de 22/01/2015 a 22/01/2016.

Inicialmente, serão coletadas as 28 variáveis de processo que mais influenciam no percentual de silício do ferro-gusa, baseado na experiência dos especialistas e operadores de alto-forno, bem como em trabalhos anteriormente desenvolvidos (BAG, 2007; CHEN, 2001; CHEN; WANG; HAN, 2010; DHOND; GUPTA; VADHAVKAR, 2000; NURKKALA; PETTERSSON; SAXÉN, 2011; SAXÉN; PETTERSSON, 2007; SAXEN; GAO, C.; GAO, Z., 2013; SINGH; SRIDHAR; DEO, 1996). Estas variáveis são responsáveis por descrever a dinâmica do processo e, por isso, são utilizadas pelos próprios operadores como parâmetros de avaliação da qualidade da produção. Também serão obtidos os dados relativos ao percentual de silício, resultante da análise laboratorial via espectrometria por raios-x. Todos os dados coletados serão tratados para a correção de *outliers* e ausência de medições, bem como para a exclusão de grandes períodos de “não-normalidade” ou paradas de alto-forno. Segundo os especialistas, entende-se por operação “normal” ou períodos de “normalidade” os momentos de regularidade do processo, isto é, instantes em que não ocorreram contingências e anomalias.

A série de silício resultante também será avaliada quanto à estacionariedade e linearidade, através de testes específicos. Também serão testados três tipos de

filtragens aplicadas aos dados, com o intuito de reduzir o nível de ruído, conforme apresentadas ao longo deste capítulo.

Após o tratamento e análise do sinal, serão avaliados e comparados diferentes modelos de previsão utilizando diversas estruturas de TDNN, tais como, NARX, NAR e NIO. Considerando informações contidas na Tabela 3 (ver Capítulo 4), a princípio o atraso máximo de cada uma das entradas será igual ao seu respectivo maior tempo médio de resposta no processo. Neste sentido, com exceção das variáveis Relação Minério/Coque e Taxa de Coque, que possuem tempo de resposta mais lento (8 horas), em todas as demais será aplicado um atraso de 3 horas.

As topologias NARX e NAR utilizam amostras autorregressivas da própria série a ser prevista. Uma vez que se deseja prever os próximos valores do teor de silício no ferro-gusa, torna-se necessário analisar a dinâmica do processo de produção para que seja possível estipular o número de atrasos de tempo os quais serão aplicados nesta série, na entrada da RNA, bem como o horizonte de previsão. Seguindo este raciocínio, para implementação das redes NARX e NAR, será estipulado um atraso de tempo de 6 horas, isto é, para prever o teor de silício das próximas amostras, serão necessários os 6 valores passados de sua entrada autorregressiva, considerando um período de amostragem de 60 minutos.

Levando-se em conta, ainda, que para as implementações das topologias NAR e NARX serão utilizados os últimos valores disponíveis das amostras de percentual de silício, para uma aplicação prática só será possível trabalhar com horizontes de previsão acima de 3 horas. Isto também é válido para a topologia NIO, que embora utilize como entradas somente os valores atuais e passados das 28 variáveis do processo, não ocorrem atualizações simultâneas de todas essas variáveis em períodos de tempo regulares, como é o caso das variáveis Temperatura do Ferro-gusa e EMF.

O tamanho da camada oculta será determinado testando-se redes com estruturas variando de 5 a 100 neurônios ocultos, incrementando 5 neurônios a cada teste. Para cada rede, será calculada a média do MSE obtido após cinco treinamentos consecutivos, utilizando o algoritmo de aprendizagem de Levenberg-Marquardt. Após todos os testes, será escolhida o tamanho da camada oculta através da rede que apresentar o menor MSE médio.

Considerando o grande número de variáveis envolvidas no processo, também serão aplicados Algoritmos de Poda (do tipo OBS), visando remover as entradas e/ou



atrasos pouco significativos para o modelo, facilitando assim, a escolha de variáveis relevantes, bem como possibilitar a redução do número de neurônios da camada oculta (NURKKALA; PETTERSSON; SAXÉN, 2011; SAXÉN; PETTERSSON, 2007). As redes neurais, bem como o Algoritmo de Poda, ambos aplicados no decorrer do trabalho tratam-se de algoritmos extraídos de um *toolbox* para Matlab, desenvolvido por Magnus Nogaard e disponibilizado para *download* no site da Universidade Técnica da Dinamarca (THE NNSYSID, 2003). Foram realizadas algumas alterações nos arquivos originais com o intuito de adaptá-los ao problema específico.

Além disto, sabendo-se que a série de silício é conhecida por apresentar características de natureza não-estacionária e diferentes componentes de frequência, este estudo também propõe a implementação de técnicas de decomposição do sinal aplicadas na análise de séries temporais. Neste sentido, será avaliada a utilização da MODWPT, como forma de decompor a série de teor de silício em subséries construtivas mais simples, representando o sinal original em bandas de frequências com diferentes resoluções, para, então, serem utilizadas como variáveis de entrada no modelo de previsão. Para a implementação da MODWPT será utilizado o *toolbox* do Matlab.

Considerando um período de amostragem de 60 minutos, para todos os testes serão selecionadas as primeiras 2.880 amostras (aproximadamente 120 dias) do conjunto de dados, divididas em dois grandes blocos. O primeiro bloco consiste das 2.160 primeiras amostras que serão utilizadas para o treinamento e validação da rede neural (1.800 amostras para o treinamento e 360 amostras para validação) e o segundo bloco, composto pelas 720 amostras (30 dias) restantes, será aplicado na validação dos testes.

Para a análise dos resultados serão utilizados critérios de desempenho como o MSE, MAPE e PEA, descritos na subseção 3.5.1. O MSE e o MAPE são medidas já consolidadas na validação de modelos de previsão (LEWIS, 1997) e aplicadas nas metodologias desenvolvidas em trabalhos anteriores (CHEN, 2001; NURKKALA; PETTERSSON; SAXÉN, 2011; SAXÉN; PETTERSSON, 2007; WANG *et al.*, 2015).

No entanto, Chen, Wang e Han (2010) e Yan-Jiong, Liang-Tu e Min (1986) afirmam que, na prática, um importante critério a ser avaliado quando se trata de modelos de previsão do teor de silício é o PEA menor que uma tolerância (variação) pré-definida. Em geral, a variação do PEA é definida de acordo com as necessidades dos

especialistas do alto-forno e, por esta razão, para este trabalho será definida uma variação de 0,05.

## 5.2 REAMOSTRAGEM DOS DADOS

Em razão da discrepância entre os períodos de amostragens das variáveis pré-selecionadas e ao interesse da siderúrgica em precisar os dados a cada uma hora, bem como, da necessidade de se trabalhar com períodos de amostragens regulares, os dados foram reamostrados de forma a possuírem valores médios horários (período de amostragem de 60 minutos).

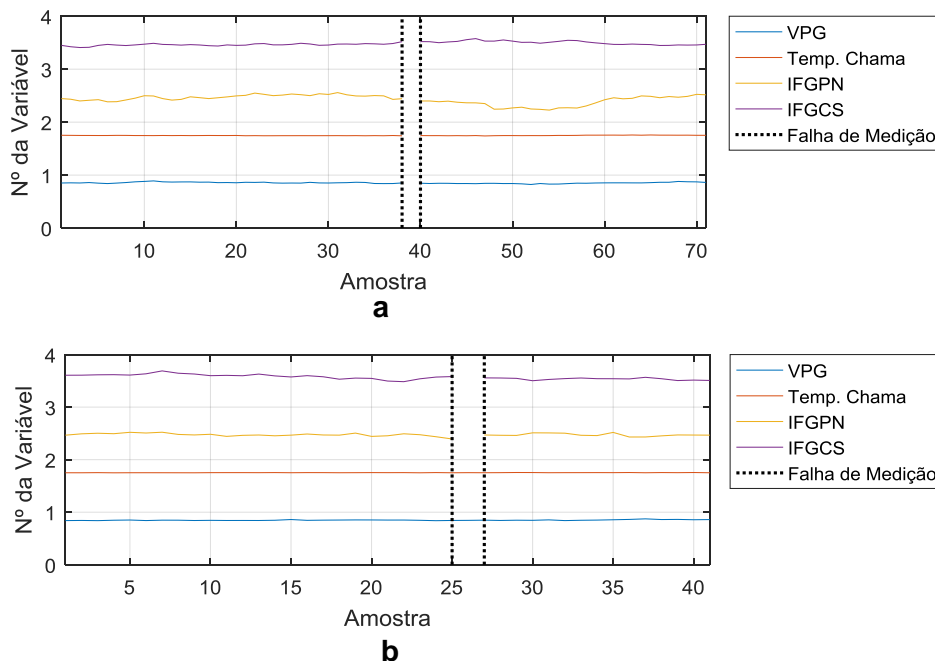
A escolha deste período de amostragem também é justificada pelo fato de que ele se aproxima ao período de amostragem da série de silício, que corresponde ao tempo de carregamento de cada carro torpedo ( $\approx 50$  minutos). Assim, para a reamostragem dos dados de processo foi utilizada a interpolação linear, uma vez que, devido à natureza do sistema, em geral, estes parâmetros não sofrem grandes variações a curto prazo.

## 5.3 TRATAMENTO DOS DADOS

Ao analisar todo o conjunto de dados, que corresponde a aproximadamente 8.760 horas de processo, foi constatada a ausência de algumas amostras. Seguindo as orientações dos especialistas, possivelmente essas ausências são justificadas por falhas dos sensores e/ou momentos de “não-normalidade” do processo.

Nos casos em que ocorreu a ausência de uma ou duas medições consecutivas, com pouca variação entre os valores das amostras que a antecedem e sucedem, estes foram considerados como falha de medição dos sensores e, por isso, foi utilizada uma interpolação linear para recuperar as informações. A Figura 26 ilustra dois exemplos de ausência de amostras de algumas das séries de VPG, Temperatura de Chama, IFGPN e IFGCS, observados durante a análise dos dados. Observa-se que, para ambos os exemplos, o processo se mantém numa operação “normal” perante esta ausência de dados.

Figura 26 – Exemplo de amostras ausentes devido as falhas de medição: a) todas as variáveis possuem amostras ausentes; b) nem todos os parâmetros possuem dados faltantes.

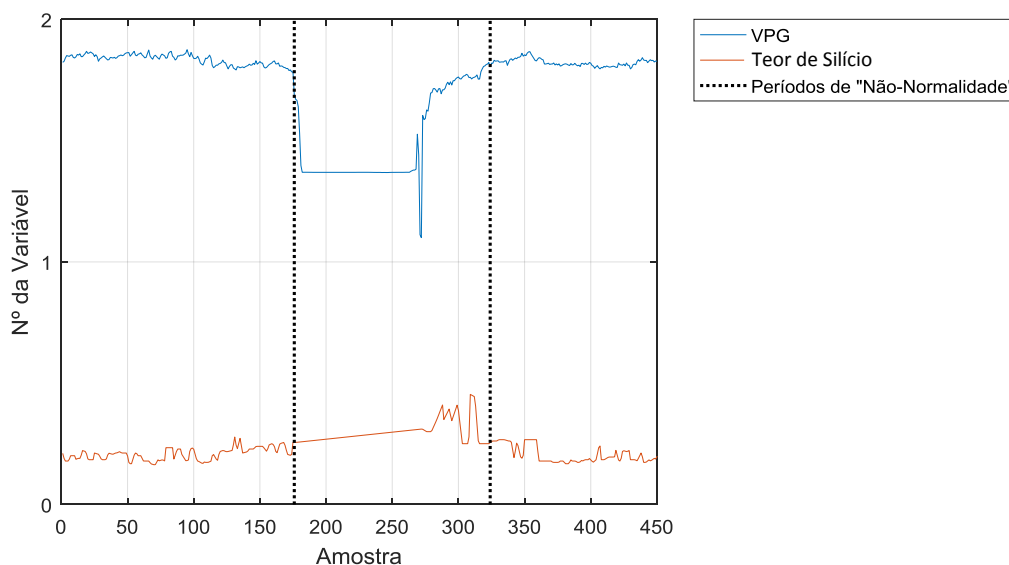


Fonte: O próprio autor.

Já durante os momentos em que ocorreram um número maior de ausência nas medições, foram considerados instantes de “não-normalidade” do processo. Outra maneira utilizada para a verificação de períodos de “não-normalidade” e paradas de alto-forno, foi através da VPG. A partir desta variável é possível mensurar o nível de produtividade do alto-forno. Em geral, valores de VPG abaixo de 4,5 ton/min, indicam uma parada de alto-forno (ou momento de “não-normalidade” do processo) e, frequentemente, estes momentos são seguidos de valores extremamente altos da série de silício. Além disto, é importante destacar que mesmo durante as paradas, alguns parâmetros continuam variando, outros são zeradas e há ainda aqueles que permanecem constantes.

A Figura 27 evidencia um exemplo de um período de “não-normalidade” identificado durante a análise das séries de dados do teor de silício no ferro-gusa e a VPG. Como esta variação na velocidade de produção é devida aos momentos de “não-normalidade” do sistema e/ou paradas de alto-forno, as amostras correspondentes aos respectivos períodos de tempo foram excluídas.

Figura 27 – Exemplo de identificação de um período de “não-normalidade”.



Fonte: O próprio autor.

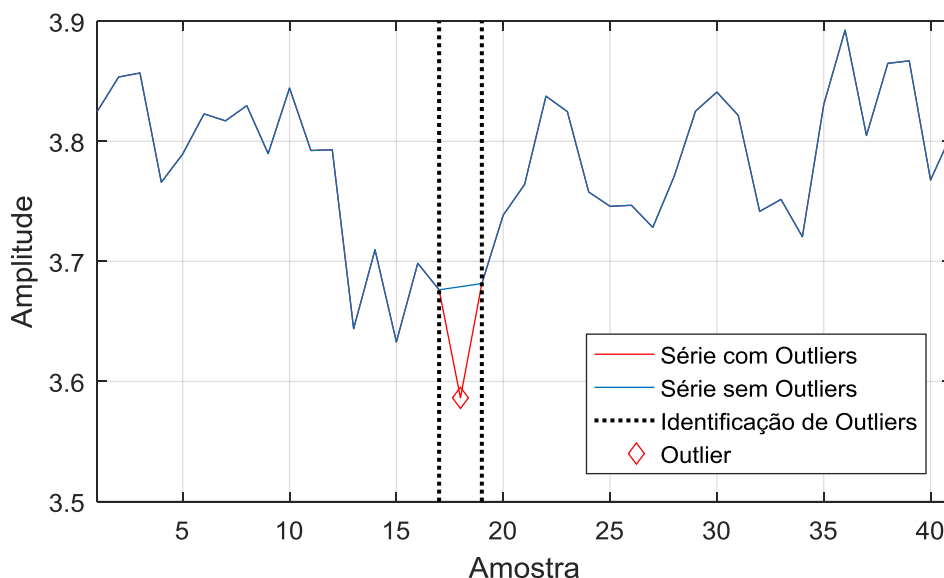
Depois de eliminados e tratados os dados referentes às falhas de medição e momentos de “não-normalidade” do processo, foi avaliada a presença de valores atípicos (*outliers*) no conjunto de dados, que podem influenciar diretamente a homogeneidade das séries. Notou-se que, em geral, os sinais de entrada não apresentavam tendência, embora apresentassem alguns valores atípicos. As ocorrências de *outliers* foram encaradas como erros de medição pois tratam-se de casos pontuais durante os instantes de “normalidade” do processo e, por isso, devem ser tratados de maneira a torná-los consistentes.

Baseando-se em Bag (2007), para a identificação dos *outliers*, foram selecionados apenas os dados que estivessem dentro do intervalo  $[\mu - 2\sigma; \mu + 2\sigma]$ . As amostras que estivessem fora do intervalo, eram eliminadas e depois interpoladas linearmente. Também foram especificados os valores máximos e mínimos assumidos por cada uma das variáveis durante uma operação “normal”. Assim, se dois desvios abaixo da média fossem menores que o mínimo, então o limite inferior do intervalo é ajustado para o valor mínimo assumido pela variável analisada. De maneira análoga, caso dois desvios acima da média fossem maiores que o máximo, então o limite superior do intervalo é ajustado para o valor máximo.

A Figura 28 apresenta um exemplo de identificação de *outliers* da variável Pressão de Sopro. O ponto em vermelho é considerado um valor atípico da variável analisada. A

série em azul é o resultado após a eliminação e interpolação do ponto identificado. Este processo é repetido ao longo das demais séries.

Figura 28 – Identificação de períodos contendo *outliers*.



Fonte: O próprio autor.

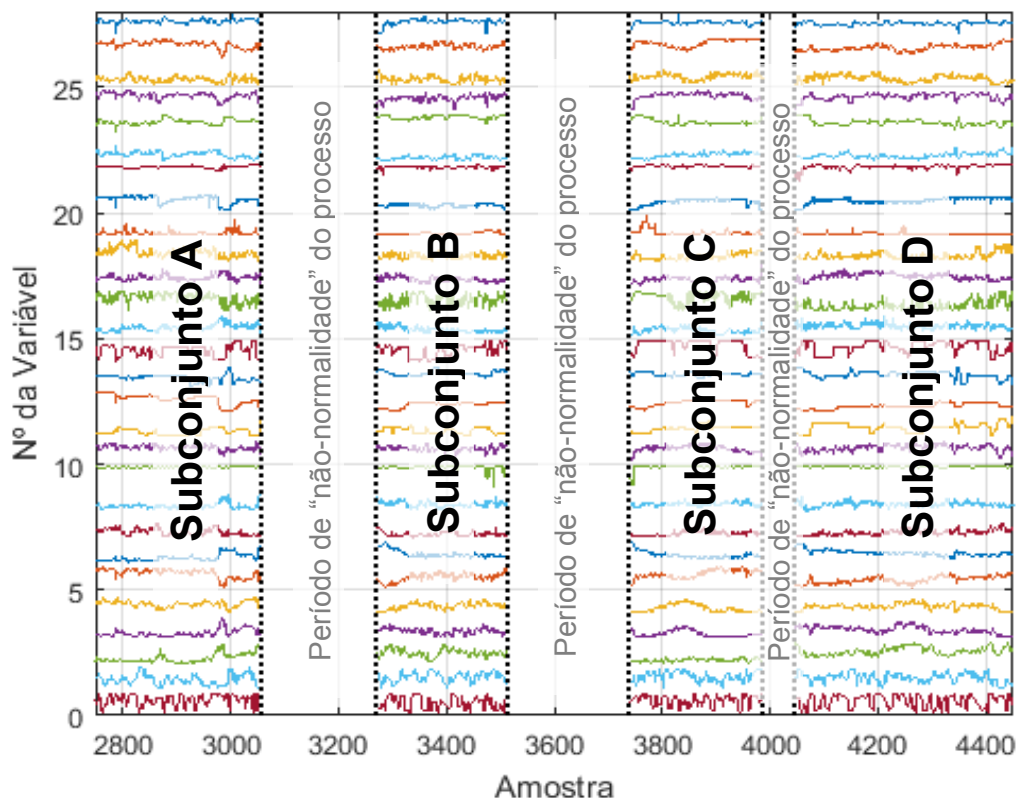
Depois de toda a análise, restaram cerca de 6.872 amostras, divididas em 10 subconjuntos. Estes subconjuntos são definidos pelos instantes em que não houveram descontinuidade do processo. Além disto, para preservar ao máximo a relação temporal entre as amostras, foram selecionados apenas os blocos com mais de 240 amostras ( $\approx 10$  dias). Deve-se frisar ainda, que dentre os fornos da ArcelorMittal, o Alto-Forno 3 é o que apresenta as maiores variações devido às condições mais difíceis de operação, justificando assim, o número de irregularidades encontradas, correspondendo a cerca de 21,55% do número total de amostras no conjunto de dados.

A Figura 29 apresenta a identificação de 4 subconjuntos de dados, para fins ilustrativos. Nesta figura, cada subconjunto é composto por um determinado número de amostras de cada uma das 28 variáveis pré-selecionadas.

Deve-se frisar que, devido ao comportamento do processo de produção do alto-forno analisado neste trabalho, as amostras do percentual de silício costumam apresentar-se fora da faixa desejada, que é de  $[0,17; 0,42]$ . Portanto, a série de silício não foi avaliada diretamente e sim, selecionados os mesmos períodos em que as demais variáveis do processo apresentaram-se dentro da “normalidade”. Neste sentido, foram

selecionadas as amostras da série de silício dos respectivos instantes abrangidos pelos subconjuntos gerados pelas variáveis pré-selecionadas.

Figura 29 – Identificação de 4 dos 10 subconjuntos de dados.



Fonte: O próprio autor.

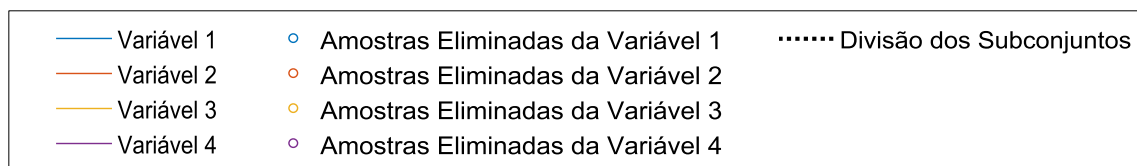
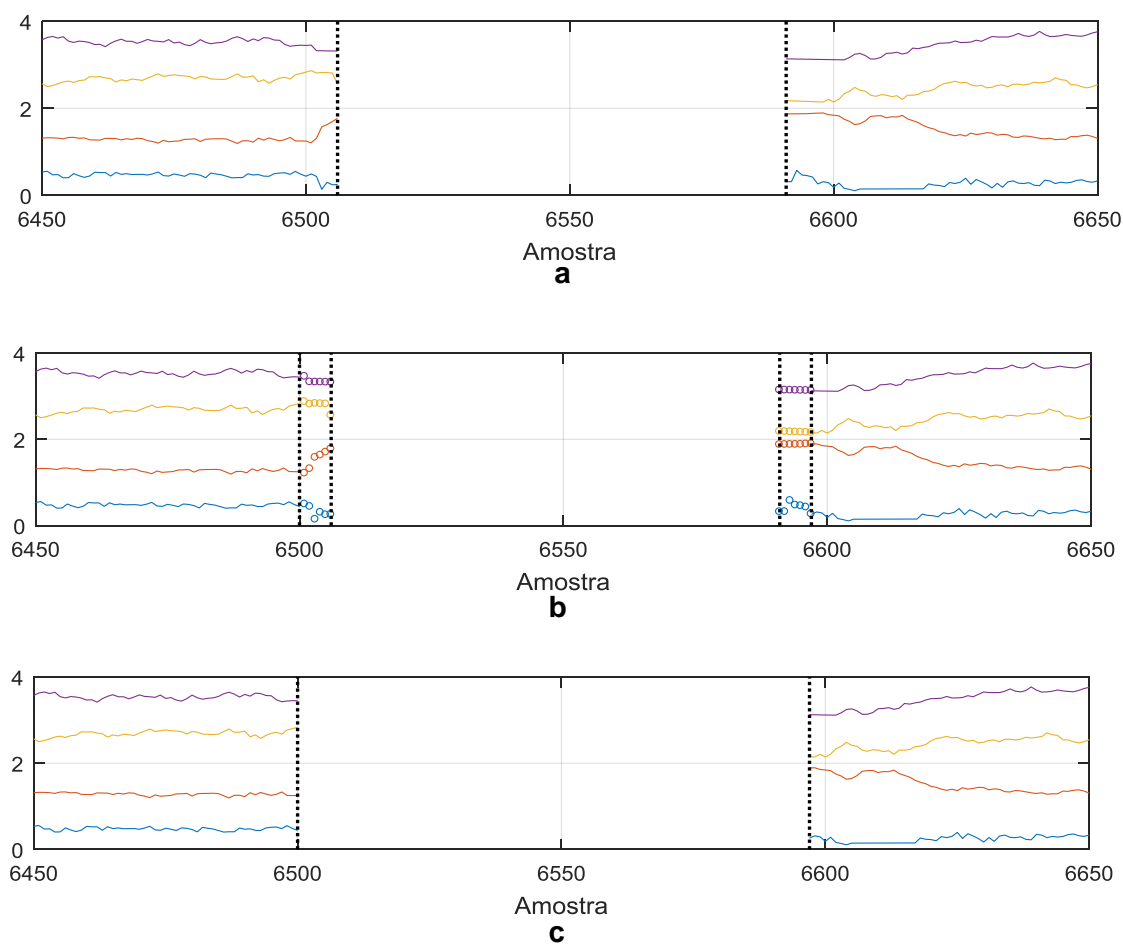
A próxima etapa do pré-processamento dos dados consiste em reagrupar os subconjuntos levando-se em conta um método que evite a descontinuidade entre eles. A primeira etapa desta união de subconjuntos, consiste na eliminação dos instantes que antecedem e sucedem períodos de “não-normalidade” do processo, excluindo-se 6 amostras no início e 6 amostras no fim de cada subconjunto. Esta eliminação visa assegurar a homogeneidade dos dados. Após esta etapa, visando evitar pontos de quebra, sem uma lógica sequencial, foram inseridas 6 amostras entre os subconjuntos, por meio da interpolação por função *spline* cúbica<sup>3</sup>.

<sup>3</sup> A *spline* cúbica é uma função polinomial por partes e contínua, com primeira e segunda derivadas contínuas. Uma vez que esta função não possui picos e nem troca abruptamente a curvatura entre os nós, a *spline* é uma técnica de interpolação conhecida por conferir maior suavidade a curva gerada entre os pontos interpolados (DE BOOR, 1978).

As Figuras 30 e 31 representam o processo de reagrupamento entre dois subconjuntos de amostras. Para melhor visualização, cada subconjunto será representado por quatro séries temporais com 200 amostras cada.

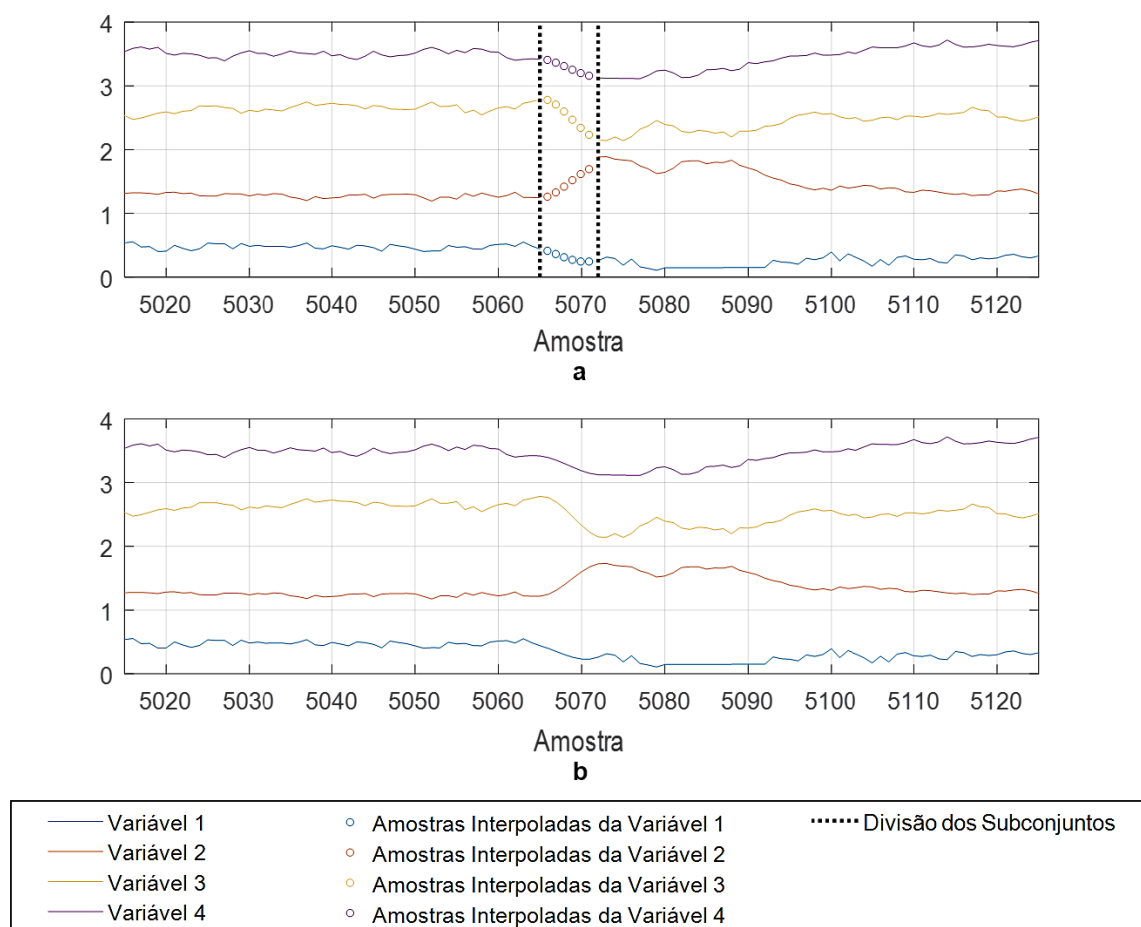
É importante ressaltar que apenas no início do primeiro subconjunto e no final do último subconjunto não foram criadas novas amostras, uma vez que, não há um subconjunto sucessor, portanto, após todas as etapas de tratamento de dados, cada série do conjunto de dados contará com 6.812 amostras.

Figura 30 – Eliminação das 6 amostras no início e no fim de cada subconjunto, representado por quatro séries temporais com 200 amostras cada: a) representação dos subconjuntos; b) evidenciadas as 6 amostras, no início e no fim de cada subconjunto, a serem eliminadas de cada variável; c) ilustrados os novos pontos de início e fim dos subconjuntos, após a eliminação das amostras.



Fonte: O próprio autor.

Figura 31 – Inserção por interpolação de 6 amostras entre os subconjuntos: a) apresentação das novas amostras inseridas em cada série através de uma interpolação; b) ilustração das respectivas séries resultantes.



Fonte: O próprio autor.

#### 5.4 NORMALIZAÇÃO DOS DADOS

Conforme evidenciado pela Tabela 3 apresentada no Capítulo 4, as variáveis de entrada apresentam uma grande diversidade de magnitudes. Sendo assim, visando garantir que todos os parâmetros possam ser comparados entre si de maneira homogênea, eles devem ser escalonados para dentro da mesma faixa de valores, através da normalização pela amplitude. Esta técnica aliada às redes neurais pode proporcionar uma redução dos erros durante o treinamento e a diminuição no tempo de processamento da rede (GE, 1999). Portanto, todas as variáveis foram normalizadas para o intervalo  $[0,1; 0,9]$ , fazendo uso da Equação (43):

$$x_{norm} = 0,1 + 0,8 \cdot \left( \frac{x_i - x_{min}}{x_{max} - x_{min}} \right) \quad (43)$$



em que  $x_i$  é o valor da variável a ser normalizada,  $x_{min}$  e  $x_{max}$  são, respectivamente, os valores mínimo e máximo assumidos pela variável  $x_i$  (ver Tabela 3) e  $x_{norm}$  é o valor normalizado desta variável (CHEN; WANG; HAN, 2010). A escolha deste intervalo propicia a não existência de valores iguais a zero na série de silício, sendo possível mais adiante, o cálculo de parâmetros de desempenho como o MAPE.

## 5.5 ANÁLISE ESTATÍSTICA DA SÉRIE DE TEOR DE SILÍCIO

Após o conjunto de dados ter sido tratado quanto a ausência de dados, erros de medição e períodos de “não-normalidade” do processo, a série de teor de silício no ferro-gusa resultante foi avaliada quanto a sua estacionariedade e linearidade, baseando-se na metodologia aplicada por Gomes (2016). Para todos os testes foi adotado um nível de significância de 5%.

Conforme apresenta a Tabela 4, por meio do teste de KPSS (KWIATKOWSKI *et al.*, 1992), é possível comprovar a não-estacionariedade da série, uma vez que a hipótese nula foi rejeitada.

Tabela 4 – Resultado do Teste de Estacionariedade.

Resultado	Possíveis resultados
$H = 0$	$H = 0$ , aceitação de $H_0$ , confirmando que a série possui tendência estacionária.
$H = 1$	$H = 1$ , rejeição de $H_0$ , em favor da hipótese alternativa de que a série possui raiz unitária.

Fonte: O próprio autor.

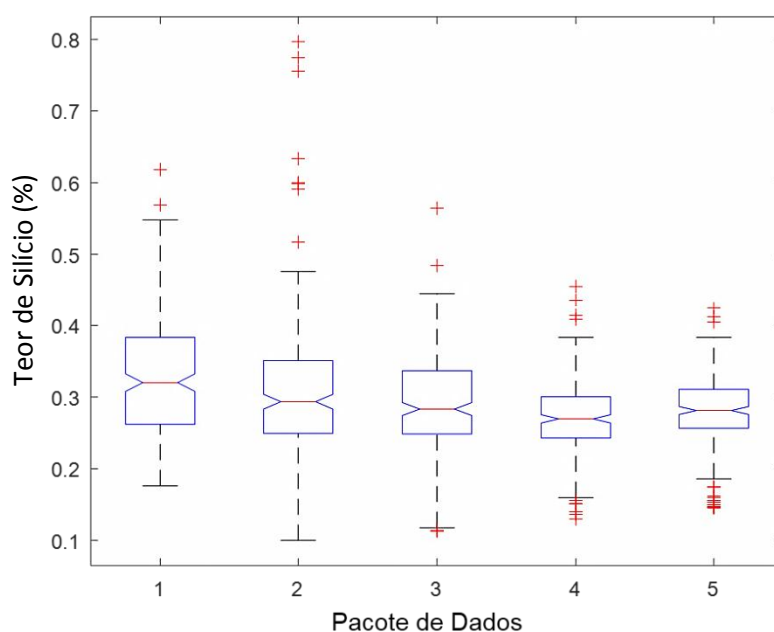
A análise da existência de não-linearidades nesta série foi realizada através dos testes de Lilliefors (LILLIEFORS, 1967) e de Kruskal-Wallis (KRUSKAL; WALLIS, 1952), cujos resultados são apresentados na Tabela 5. Para a implementação deste último teste, foi necessário dividir os dados em cinco pacotes, contendo 10 dias de dados em cada pacote, visando verificar se a média e a variância destes dados permanecia constante. Conforme ilustra a Figura 32, é possível observar graficamente que amostras não provêm da mesma distribuição.

Tabela 5 – Resultados dos Testes de Linearidade

Teste	Resultado	Possíveis resultados
Lilliefors	1	$H = 0$ : aceitação de $H_0$ , confirmando que a série é normalmente distribuída. $H = 1$ : rejeição de $H_0$ , em favor da hipótese alternativa de que a série não é normalmente distribuída.
Kruskal-Wallis	$9.9875 \times 10^{-14}$	$p$ -valor $\approx 0$ : as amostras não provêm da mesma distribuição. $p$ -valor $\approx 1$ : as amostras provêm da mesma distribuição.

Fonte: O próprio autor.

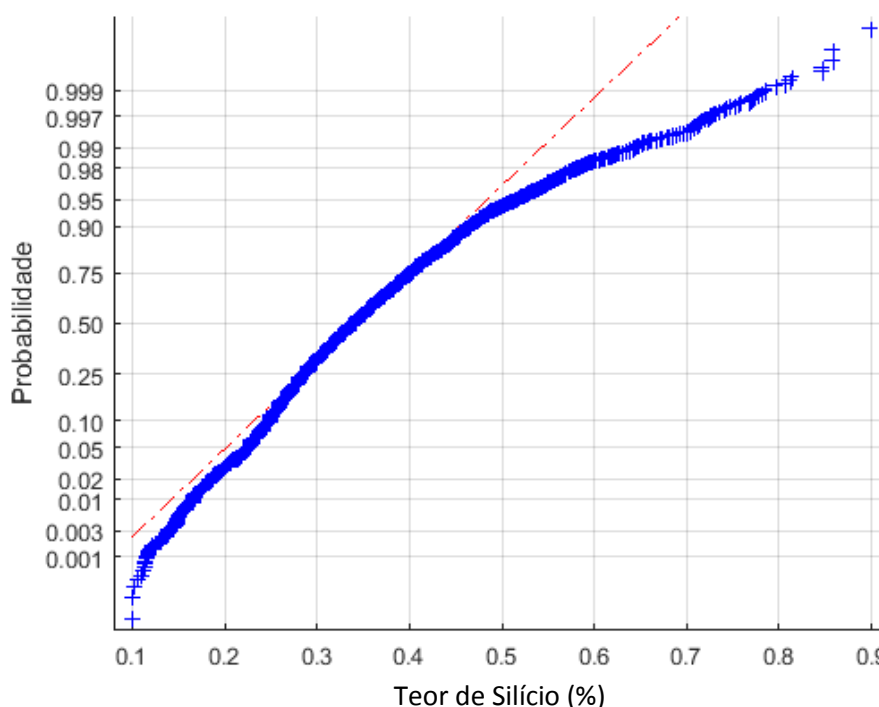
Figura 32 – Comparação entre médias e variâncias de cinco pacotes de dados da série de conteúdo de silício.



Fonte: O próprio autor.

Com base nestas análises conclui-se que a série temporal possui componentes não-lineares. Isto é confirmado pela Figura 33, onde se nota que a série de teor de silício apresenta um comportamento não-linear concentrado nos valores próximos às bordas.

Figura 33 – Comportamento da série de Conteúdo de Silício em relação à linearidade.



Fonte: O próprio autor.

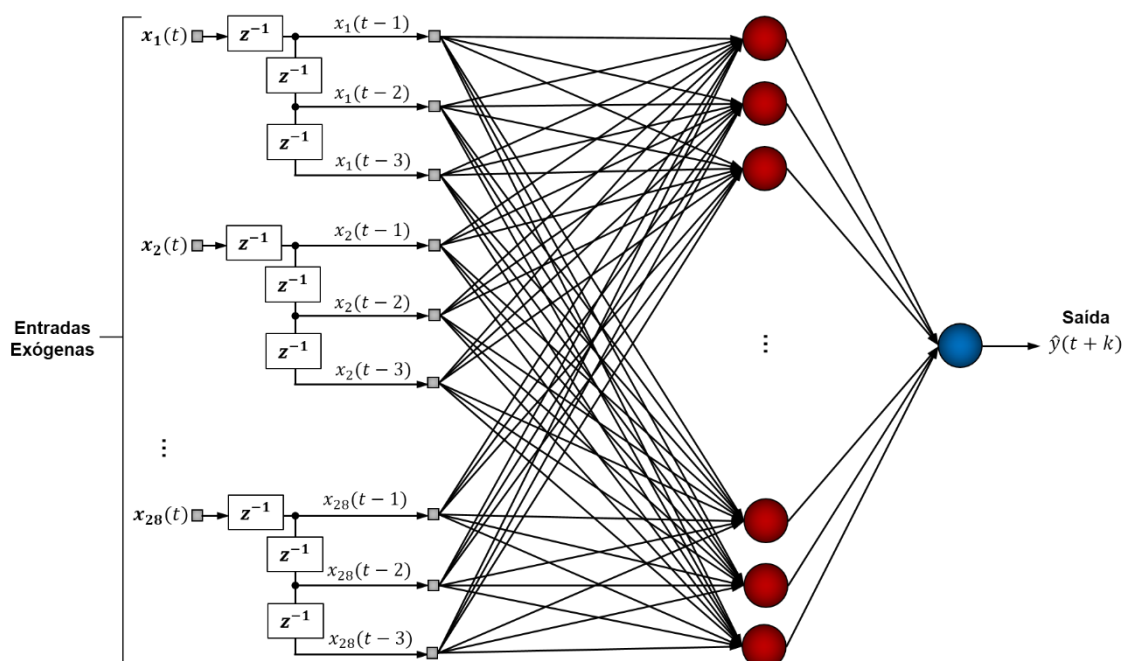
## 5.6 MODELAGEM E PREVISÃO DO TEOR DE SILÍCIO UTILIZANDO RNA

Nesta seção serão testadas variações de três topologias de RNA com atraso de tempo: NIO, NARX e NAR. Inicialmente, nas topologias NIO e NARX, nas quais são aplicadas entradas exógenas, serão utilizadas as 28 variáveis pré-selecionadas e, conforme descrito anteriormente, o Algoritmo de Poda, baseado na Matriz Hessiana e utilizando o método OBS, será o responsável por apurar quais destas variáveis apresentam maior grau de significância ao modelo. Devido à natureza deste algoritmo, os atrasos de cada entrada também serão definidos através do mesmo.

Através das Figuras 34, 35 e 36, respectivamente, é possível visualizar as três topologias de redes neurais empregadas neste trabalho: NIO, NAR e NARX. Todas as redes possuirão uma camada de entrada (com tamanho variável, de acordo com a quantidade de entradas empregadas), uma camada oculta e uma camada de saída (com apenas um neurônio, representando a única saída das redes). Os tamanhos da camada oculta, para cada uma das redes, foram obtidos pelo teste descrito na Seção 5.1, sendo 95 neurônios ocultos para rede NIO, 90 para rede NARX e 100 para rede NAR (ver Figura 37). Além disto, foi utilizada uma função de ativação do tipo tangente

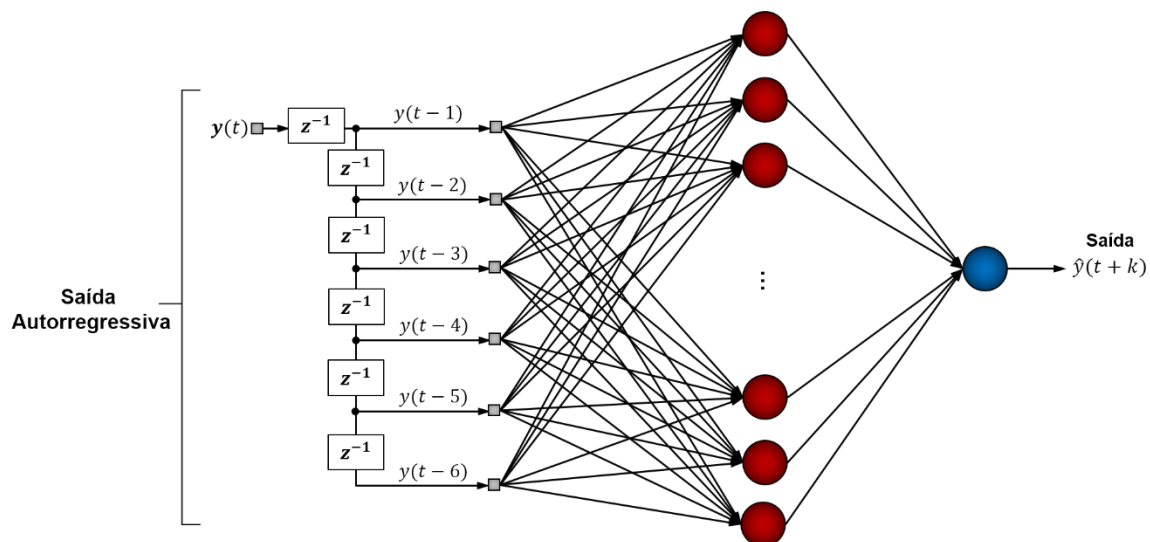
hiperbólica para os neurônios da camada oculta e uma função de ativação linear para o neurônio da camada de saída.

Figura 34 – Estrutura da topologia de rede NIO aplicada neste trabalho.



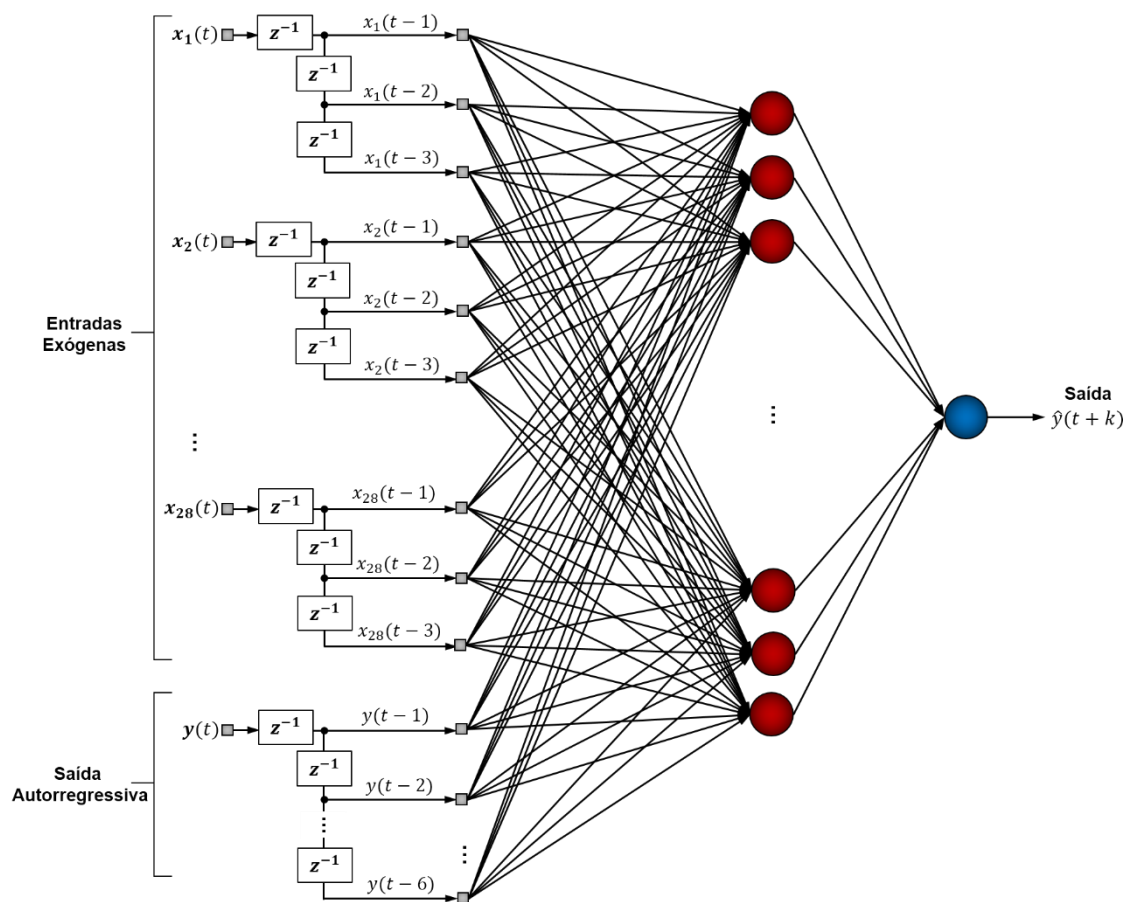
Fonte: O próprio autor.

Figura 35 – Estrutura da topologia de rede NAR aplicada neste trabalho.



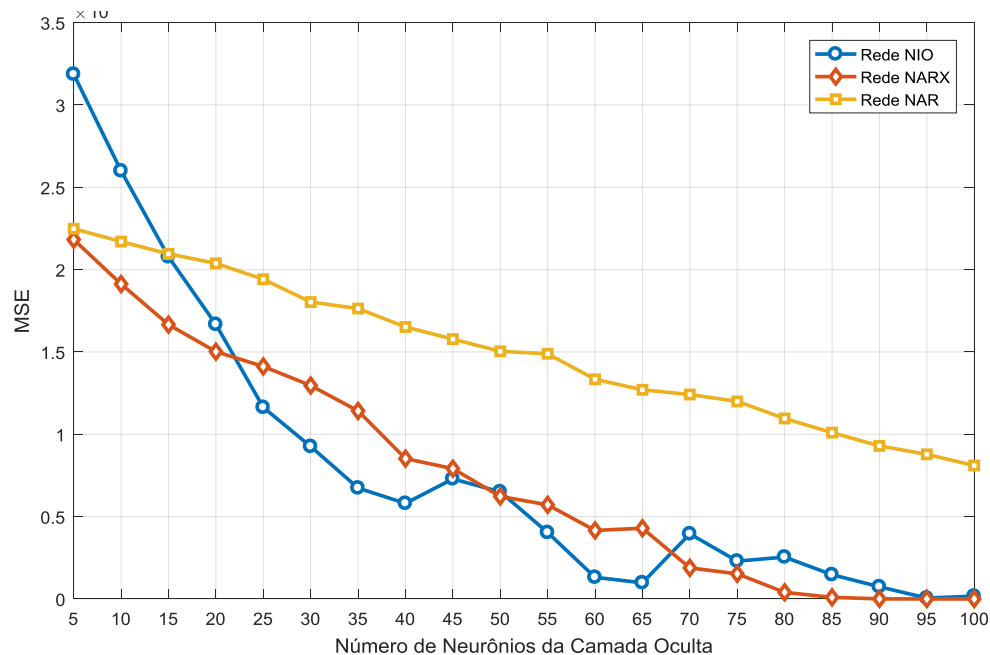
Fonte: O próprio autor.

Figura 36 – Estrutura da topologia de rede NARX aplicada neste trabalho.



Fonte: O próprio autor.

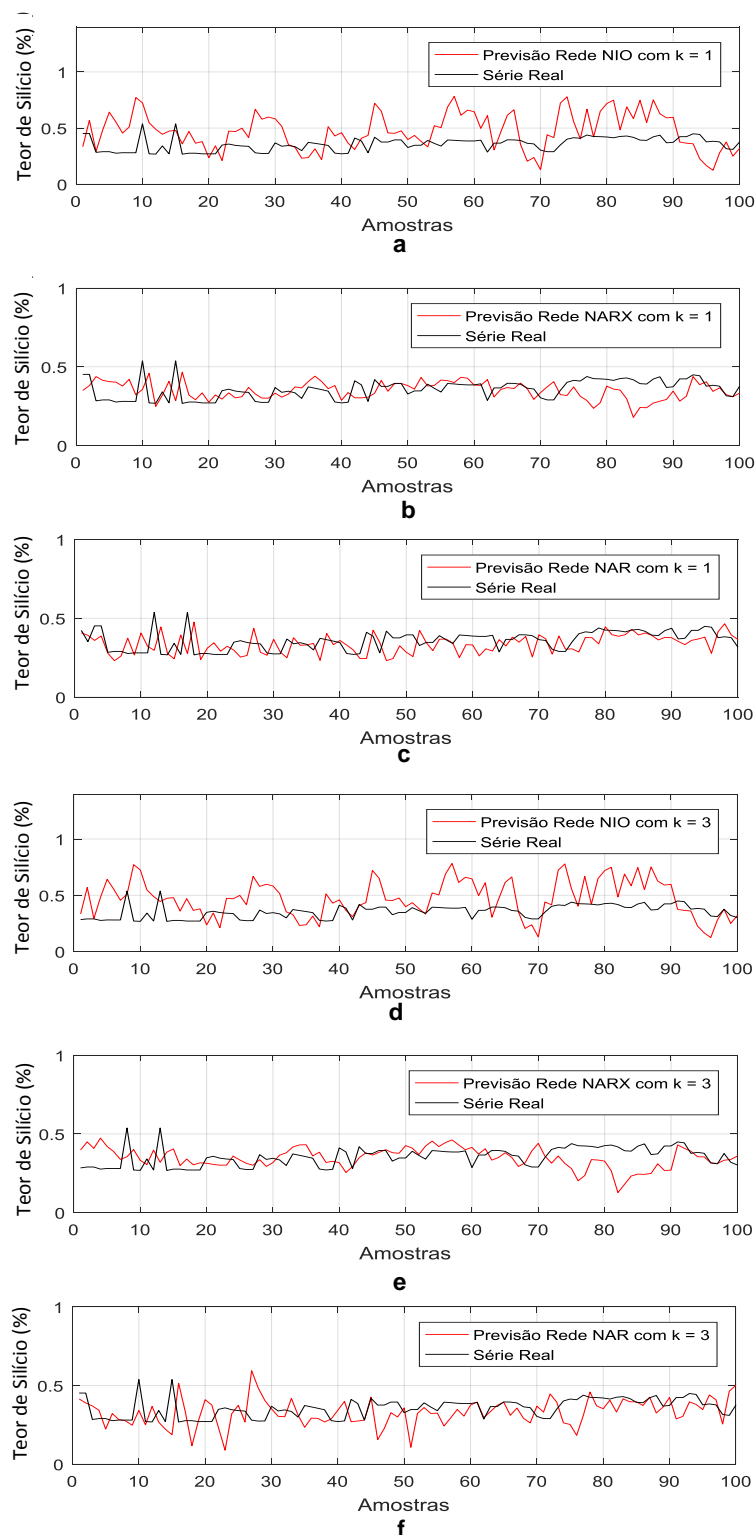
Figura 37 – O MSE médio de treinamento obtido para cada tamanho de camada oculta.



Fonte: O próprio autor.

As previsões de 1 e 3 passos à frente do teor de silício no ferro-gusa obtidas por cada uma das topologias de rede podem ser vistas através da Figura 38.

Figura 38 – Previsão do Conteúdo de Silício do ferro-gusa: a) Rede NIO com horizonte de previsão de 1 hora; b) Rede NARX com horizonte de previsão de 1 hora; c) Rede NAR com horizonte de previsão de 1 hora; d) Rede NIO com horizonte de previsão de 3 horas; e) Rede NARX com horizonte de previsão de 3 horas; f) Rede NAR com horizonte de previsão de 3 horas.



Fonte: O próprio autor.

A partir da Figura 38, nota-se que, para os modelos NAR e NARX, conforme aumenta-se o horizonte de previsão, maiores são as diferenças entre as amostras da saída prevista e as amostras do sinal original. Observa-se também que, as previsões as quais são empregadas o modelo NIO, não conseguem acompanhar as variações do sinal original.

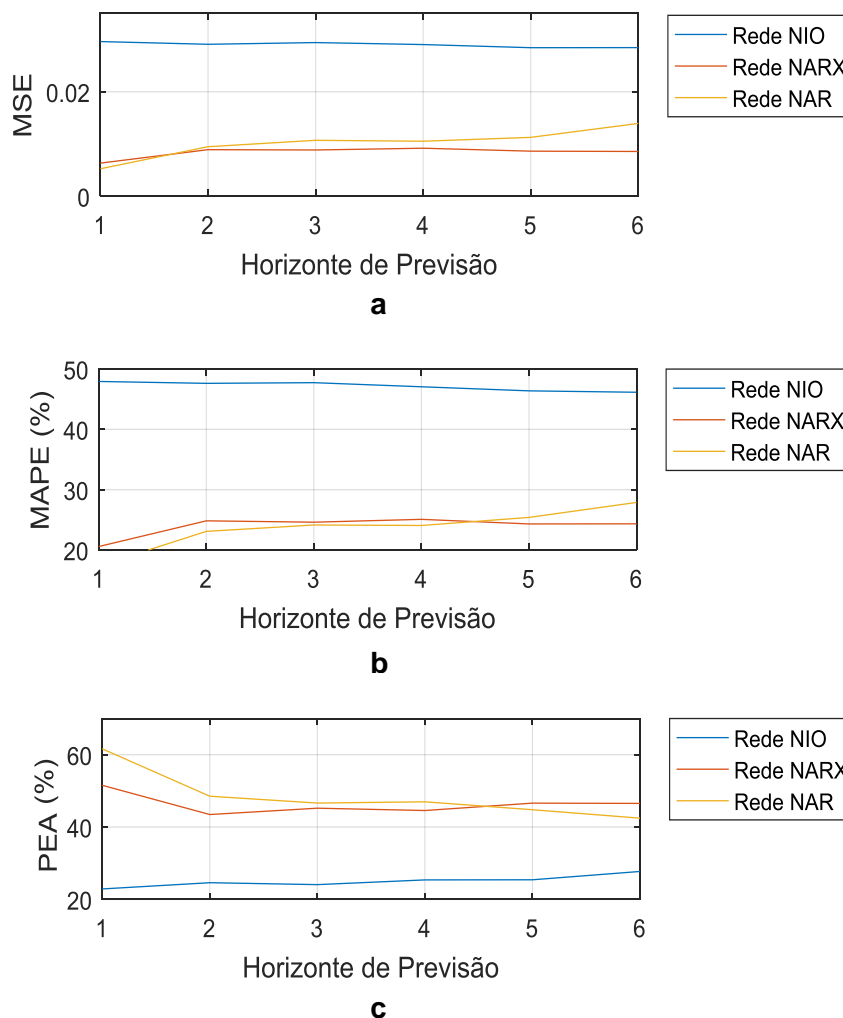
A Tabela 6 apresenta os desempenhos obtidos durante as previsões de 1 a 6 passos à frente para cada uma das topologias de rede. Os mesmos resultados também são representados pela Figura 39. É possível averiguar que, embora tenha apresentado resultados menos satisfatórios, as previsões do modelo NIO foram as que apresentam as menores deteriorações conforme aumentou-se o horizonte de previsão.

Tabela 6 – Critérios de desempenho obtidos pelas previsões.

<b>Modelo</b>	<b>Horizonte de Previsão</b>	<b>MSE</b>	<b>MAPE (%)</b>	<b>PEA (%)</b>
NIO	1 hora	0,029598	47,946	22,89326
	2 horas	0,02908	47,633	24,61322
	3 horas	0,029407	47,744	24,08451
	4 horas	0,029028	47,065	25,38787
	5 horas	0,028428	46,387	25,42373
	6 horas	0,028442	46,155	27,72277
NARX	1 hora	0,00635	20,569	51,54494
	2 horas	0,008923	24,831	43,45992
	3 horas	0,008861	24,606	45,21127
	4 horas	0,009199	25,079	44,56982
	5 horas	0,008633	24,317	46,61017
	6 horas	0,008576	24,335	46,53465
NAR	1 hora	0,005274	17,35	61,62465
	2 horas	0,009486	23,083	48,52735
	3 horas	0,010711	24,142	46,62921
	4 horas	0,010534	24,062	46,97609
	5 horas	0,011277	25,398	44,78873
	6 horas	0,013938	27,883	42,45416

Fonte: O próprio autor.

Figura 39 – Critérios de Desempenho versus Horizonte de Previsão: a) MSE; b) MAPE (%); c) PEA (%).



Fonte: O próprio autor.

Observa-se também que, em geral, menos da metade das previsões estão dentro da faixa de PEA. A ocorrência destes erros de previsão pode ser devido aos ruídos inerentes das medições das variáveis que, frequentemente, podem ser atenuados através do uso de filtros. Para tanto, foram avaliados filtros lineares mais robustos, tais como os consolidados filtros Média Móvel e Mediana Móvel (OPPENHEIM; SCHAFER; BUCK, 1999), bem como filtros não-lineares, através das *Wavelets*. É interessante destacar que, após a decomposição *Wavelet*, a maior parte das informações referentes aos ruídos são encontradas nas subséries de alta frequência. Em geral, estas subséries apresentam pouca informação a respeito do sinal original. Assim, é possível ajustar o sinal resultante de maneira a conter maior nível de



informação de baixa frequência e pouca informação de alta frequência (MISITI *et al.*, 2017).

As Tabelas 7, 8 e 9 apresentam os desempenhos obtidos durante as previsões de 1 a 6 passos à frente, respectivamente, para as topologias de rede NIO, NARX e NAR, empregando três filtros distintos, aplicando os mesmos tamanhos de camada oculta, entradas e atrasos, utilizados no teste anterior. Para os filtros de Média Móvel e Mediana Móvel, foi aplicada uma janela móvel de 3 amostras, sendo possível obter um sinal filtrado com menos variações abruptas e, ao mesmo tempo, que conseguisse acompanhar adequadamente o sinal original. Já para o filtro através das *Wavelets*, foi utilizada uma *wavelet* do tipo Symlet de ordem 4, por assemelhar-se mais com a forma característica do sinal coletado.

Tabela 7 – Comparativo dos critérios de desempenho das previsões aplicando Filtros de Média Móvel, Mediana Móvel e *Wavelet* aos dados da Rede NIO.

<b>Tipo de Filtro</b>	<b>Horizonte de Previsão</b>	<b>MSE</b>	<b>MAPE (%)</b>	<b>PEA (%)</b>
MÉDIA MÓVEL	1 hora	0,011317	29,792	38,061
	2 horas	0,011508	29,821	39,099
	3 horas	0,011823	30,071	37,887
	4 horas	0,012161	30,295	36,389
	5 horas	0,012367	30,544	36,440
	6 horas	0,01235	30,655	35,360
MEDIANA MÓVEL	1 hora	0,012164	28,168	38,483
	2 horas	0,012551	28,792	36,427
	3 horas	0,012795	29,079	35,915
	4 horas	0,012856	29,360	34,978
	5 horas	0,012873	29,457	35,875
	6 horas	0,0129	29,860	35,360
WAVELET	1 hora	0,02178	39,886	28,651
	2 horas	0,021635	39,250	28,410
	3 horas	0,021158	38,853	27,323
	4 horas	0,020544	38,282	30,606
	5 horas	0,020298	38,052	30,932
	6 horas	0,020483	38,254	28,147

Fonte: O próprio autor.

Tabela 8 – Comparativo dos critérios de desempenho das previsões aplicando Filtros de Média Móvel, Mediana Móvel e *Wavelet* aos dados da Rede NARX.

Tipo de Filtro	Horizonte de Previsão	MSE	MAPE (%)	PEA (%)
MÉDIA MÓVEL	1 hora	0,007448	21,350	50,421
	2 horas	0,014012	29,581	35,583
	3 horas	0,018402	34,627	30,000
	4 horas	0,017249	33,381	31,734
	5 horas	0,018716	34,818	30,790
	6 horas	0,019895	35,648	29,985
MEDIANA MÓVEL	1 hora	0,027795	45,119	23,595
	2 horas	0,042546	55,783	17,862
	3 horas	0,03969	53,804	19,295
	4 horas	0,030714	46,802	24,964
	5 horas	0,030195	47,219	22,175
	6 horas	0,033124	49,253	21,782
WAVELET	1 hora	0,004186	17,581	55,617
	2 horas	0,015171	34,411	30,942
	3 horas	0,01342	32,157	33,802
	4 horas	0,011652	29,555	36,530
	5 horas	0,012374	30,429	34,745
	6 horas	0,011875	29,045	36,633

Fonte: O próprio autor.

Tabela 9 – Comparativo dos critérios de desempenho das previsões aplicando Filtros de Média Móvel, Mediana Móvel e *Wavelet* aos dados da Rede NAR (Continua).

Tipo de Filtro	Horizonte de Previsão	MSE	MAPE (%)	PEA (%)
MÉDIA MÓVEL	1 hora	0,000627	5,831	94,678
	2 horas	0,002113	10,845	78,541
	3 horas	0,004065	15,017	65,730
	4 horas	0,005034	16,872	60,478
	5 horas	0,005519	17,918	55,775
	6 horas	0,005976	18,719	55,430

Tabela 9 – Comparativo dos critérios de desempenho das previsões aplicando Filtros de Média Móvel, Mediana Móvel e *Wavelet* aos dados da Rede NAR (Conclusão).

Tipo de Filtro	Horizonte de Previsão	MSE	MAPE (%)	PEA (%)
MEDIANA MÓVEL	1 hora	0,001819	9,873	80,252
	2 horas	0,004614	16,244	59,748
	3 horas	0,006903	20,244	50,000
	4 horas	0,007696	21,667	48,945
	5 horas	0,00773	21,561	49,859
	6 horas	0,008071	22,094	46,544
WAVELET	1 hora	0,000257	3,343	98,179
	2 horas	0,001221	7,937	87,237
	3 horas	0,003057	12,883	73,736
	4 horas	0,004406	15,079	66,948
	5 horas	0,009075	21,765	53,803
	6 horas	0,015177	29,505	40,480

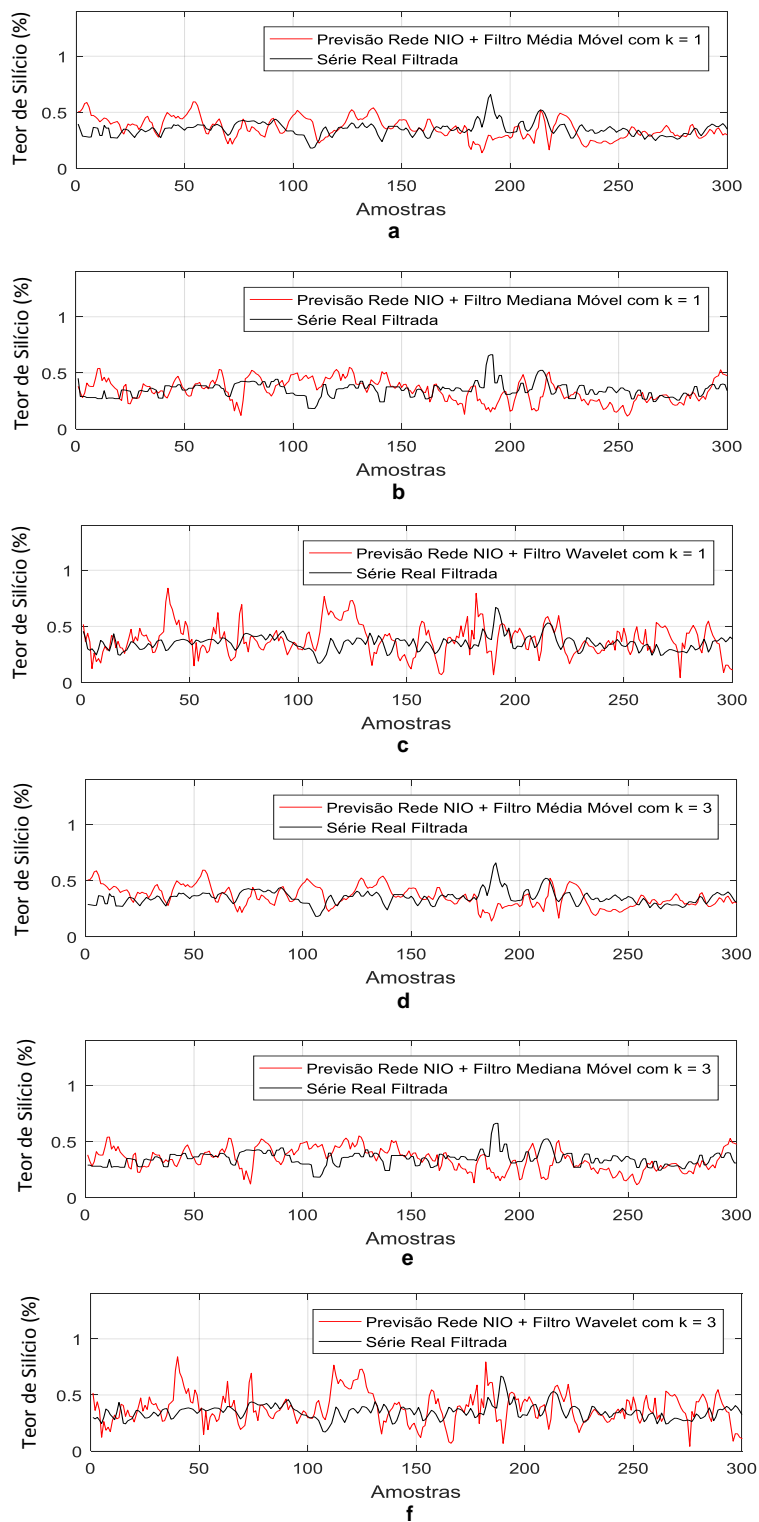
Fonte: O próprio autor.

Com exceção da Rede NARX, nota-se uma melhoria significativa no desempenho das redes, sobretudo, as quais empregaram os filtros de Média Móvel e *Wavelet*. Destaca-se a obtenção de um PEA de 98,179% para o modelo NAR (horizonte de previsão de 1 hora), ao serem implantados os filtros *Wavelet*. Contudo, ao observar as previsões obtidas pela topologia NAR, ilustradas pela Figura 42, é possível constatar que pequenos atrasos pertinentes às previsões ainda ocorreram.

Um problema relacionado às técnicas que utilizam a janela móvel é que, a cada nova amostra resultante, não são capturadas muitas novas informações. Em vez disso, cada novo conjunto de 3 amostras, possui apenas um novo ponto de dados comparado ao conjunto anterior. O conjunto de dados resultante da janela móvel será muito mais suave devido aos pontos de dados sobrepostos.

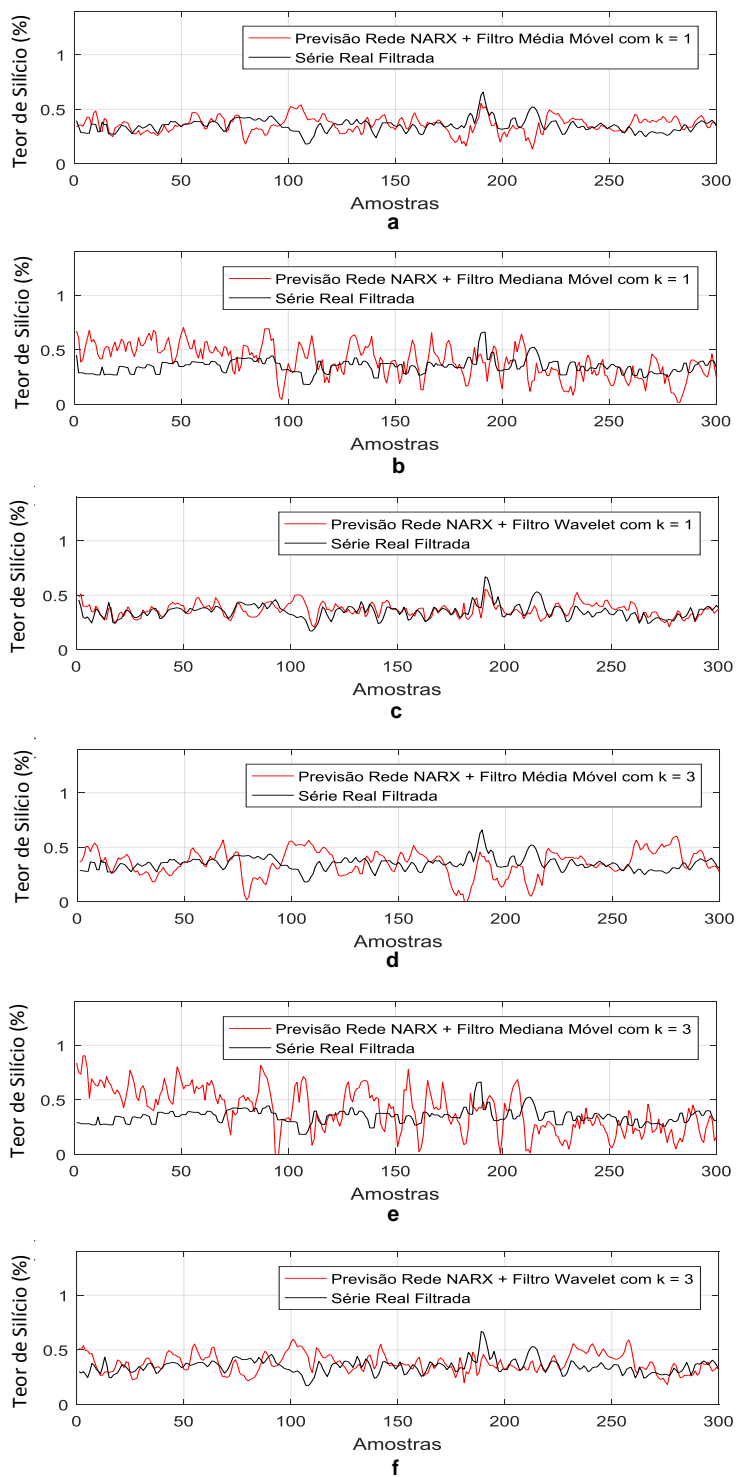
A aplicação do filtro Média Móvel nos dados de entrada das redes neurais, por exemplo, faz com que a curva do valor teor de silício resultante seja mais suave do que a curva dos valores médios horários. Em seu estudo, Dhond (2000) sugere que esta característica dos filtros utilizando técnicas de janelamento faz com a RNA seja capaz de prever os valores da curva mais suave com maior facilidade, porque a rede terá maior dependência com as amostras anteriores da série de silício.

Figura 40 – Previsão do Conteúdo de Silício do ferro-gusa empregando uma Rede NIO e Filtro a) Média Móvel e horizonte de previsão de 1 hora; b) Mediana Móvel e horizonte de previsão de 1 hora; c) *Wavelet* e horizonte de previsão de 1 hora; d) Média Móvel e horizonte de previsão de 3 horas; e) Mediana Móvel e horizonte de previsão de 3 horas; f) *Wavelet* e horizonte de previsão de 3 horas.



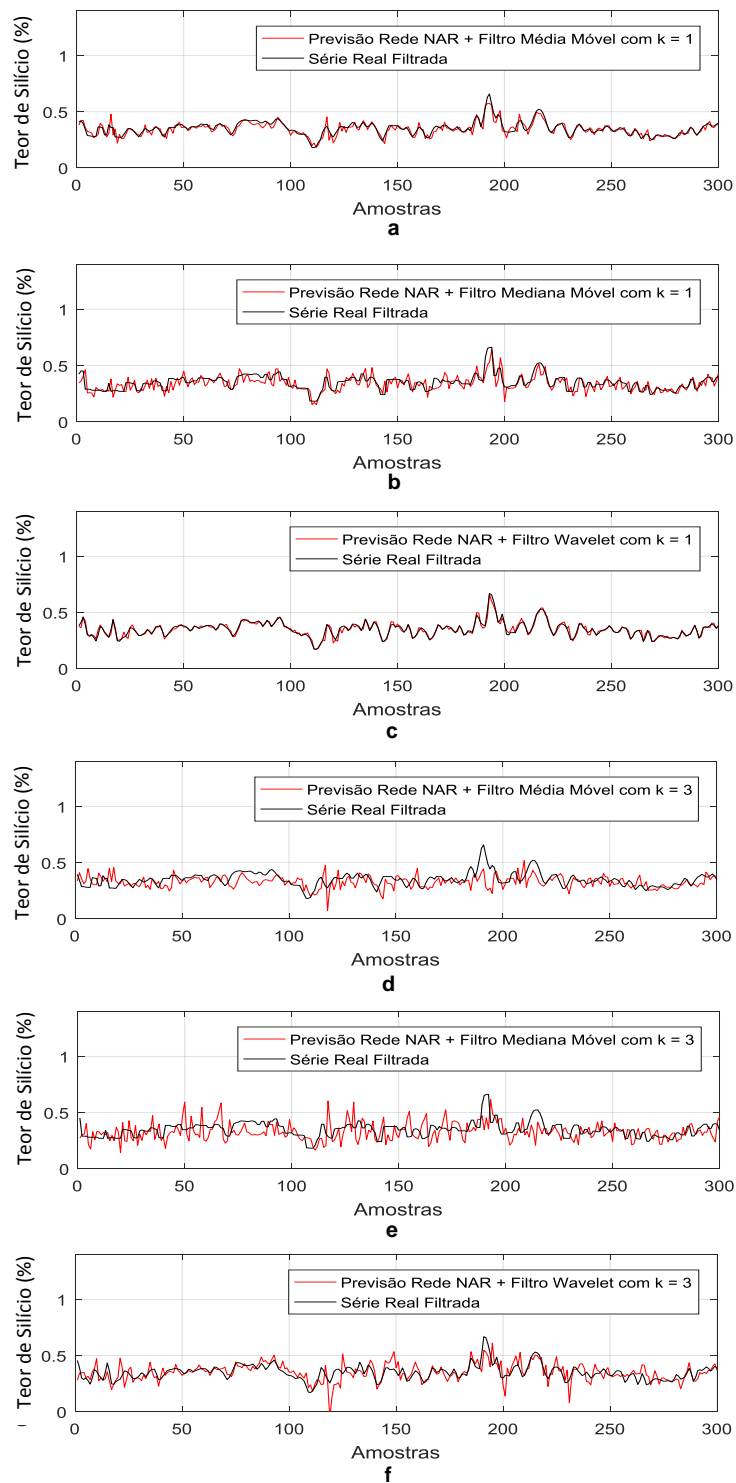
Fonte: O próprio autor.

Figura 41 – Previsão do Conteúdo de Silício do ferro-gusa empregando uma Rede NARX e Filtro a) Média Móvel e horizonte de previsão de 1 hora; b) Mediana Móvel e horizonte de previsão de 1 hora; c) *Wavelet* e horizonte de previsão de 1 hora; d) Média Móvel e horizonte de previsão de 3 horas; e) Mediana Móvel e horizonte de previsão de 3 horas; f) *Wavelet* e horizonte de previsão de 3 horas.



Fonte: O próprio autor.

Figura 42 – Previsão do Conteúdo de Silício do ferro-gusa empregando uma Rede NAR e Filtro a) Média Móvel e horizonte de previsão de 1 hora; b) Mediana Móvel e horizonte de previsão de 1 hora; c) *Wavelet* e horizonte de previsão de 1 hora; d) Média Móvel e horizonte de previsão de 3 horas; e) Mediana Móvel e horizonte de previsão de 3 horas; f) *Wavelet* e horizonte de previsão de 3 horas.



Fonte: O próprio autor.

É possível notar através das Figuras 40, 41 e 42 que a aplicação do filtro *Wavelet* nos dados de entrada da rede NAR foi a combinação que apresentou as melhores previsões, embora que, para horizontes de previsão maiores ocorreram deterioração do desempenho. Além disto, com relação as topologias NIO e NARX, em virtude do grande número de variáveis exógenas empregadas na camada de entrada, o excesso de informações e consequente redundância dos dados podem confundir a rede durante o processo de aprendizagem. Isto também pode ser justificado pelo longo tempo transcorrido para a conclusão do treinamento (cerca de 75 minutos em um PC Padrão, Core i3 2,4GHz).

Com o intuito de eliminar a redundância dos dados e, consequentemente, aumentar o desempenho da previsão, empregou-se o Algoritmo de Poda do tipo OBS nas topologias de redes. Considerando ainda que, de uma maneira geral, as previsões as quais empregavam filtros *Wavelet* foram as que apresentaram os resultados mais satisfatórios, esta implementação será mantida durante a aplicação da poda. É importante mensurar que em virtude da complexidade do OBS e do consequente longo tempo de processamento, as redes foram inicializadas com 10 neurônios ocultos.

No entanto, para a avaliação das entradas mais significativas resultantes do OBS, o Algoritmo de Poda foi implementado 10 vezes. No final de cada implementação (poda), as entradas remanescentes, bem como os seus respectivos atrasos, foram armazenadas. Nos Anexos B e C é possível verificar as entradas com as respectivas ordens dos regressores<sup>4</sup> remanescentes no final de cada implementação do Algoritmo de Poda, bem como os critérios de desempenho obtidos para uma previsão de 1 passo à frente.

É possível organizar estes resultados de maneira a representar a ocorrência dos respectivos atrasos de cada uma das entradas, conforme mostra a Tabela 10. Nesta análise, cada uma das conexões entre um neurônio da camada oculta e a respectiva entrada/atraso serão computadas separadamente. Suponha, por exemplo, que o regressor de ordem 3 para uma determinada entrada apresente conexões com 3 neurônios da camada oculta, neste caso, serão consideradas 3 ocorrências do conjunto entrada/atraso.

---

<sup>4</sup> Entende-se por ordem dos regressores, os atrasos de tempo os quais são deslocados cada uma das entradas do modelo (AGUIRRE, 2015).

Dos resultados apresentados pela Tabela 10, referentes às podas aplicadas à rede NIO, é possível identificar a ocorrência com maior frequência de algumas entradas como Taxa de Coque, *PCI Rate*, Temperatura do Ferro-gusa e Taxa de O<sub>2</sub>.

Tabela 10 – Ocorrência das entradas e ordem dos regressores resultantes após cada implementação do Algoritmo de Poda aplicado às Rede NIO e NARX (Continua).

REDE NIO		REDE NARX	
Variável de Entrada	Somatório das Ocorrências	Variável de Entrada	Somatório das Ocorrências
Taxa de Coque	23	Saída Autorregressiva	116
Temperatura do ferro-gusa	17	Taxa de Coque	68
<i>PCI Rate</i>	14	Relação Minério/Coque	60
Taxa de O <sub>2</sub>	14	Temperatura de Sopro	44
IFGCN	9	IFGPN	38
Força-Eletromotriz	9	IFGCS	32
Relação CO/CO <sub>2</sub>	8	Índice térmico H0	31
Relação Minério/Coque	8	Temperatura do gás de topo	31
Teor de CO <sub>2</sub> no topo	8	IFGPS	30
Temperatura de Chama	7	Força-Eletromotriz	30
IFGCS	7	Temperatura do ferro-gusa	29
Volume de Sopro	7	IFGCN	28
Temperatura de Sopro	7	Eficiência do gás H <sub>2</sub>	27
Pressão de Sopro	7	Redução Direta	26
Eficiência do gás H <sub>2</sub>	6	Volume de Sopro	26
IFGPN	6	Temperatura de Chama	25
IFGPS	6	Teor de H <sub>2</sub> no topo	25
Umidade de Sopro	6	Velocidade de Produção de Gusa	24



Tabela 10 – Ocorrência das entradas e ordem dos regressores resultantes após cada implementação do Algoritmo de Poda aplicado às Rede NIO e NARX (Conclusão).

REDE NIO		REDE NARX	
Variável de Entrada	Somatório das Ocorrências	Variável de Entrada	Somatório das Ocorrências
Temperatura do gás de topo	6	Perdas Térmicas pelo <i>Stave</i>	24
Teor de CO no topo	6	Pressão de topo	23
Velocidade de Produção de Gusa	5	Eficiência do gás CO	23
Redução Direta	5	<i>PCI Rate</i>	22
Teor de N <sub>2</sub> no topo	5	Umidade de Sopro	22
Eficiência do gás CO	5	Teor de CO <sub>2</sub> no topo	21
Perdas Térmicas pelo <i>Stave</i>	5	Taxa de O <sub>2</sub>	19
Teor de H <sub>2</sub> no topo	3	Relação CO/CO <sub>2</sub>	18
Índice térmico H0	1	Teor de CO no topo	18
Pressão de topo	1	Teor de N <sub>2</sub> no topo	14

Fonte: O próprio autor.

Conforme os dados do Anexo B, os resultados obtidos pela nona implementação do Algoritmo de Poda resultou em uma rede com apenas três entradas (Taxa de Coque, Temperatura do ferro-gusa e Taxa de O<sub>2</sub>) e MSE de 0,0051, MAPE de 19,10% e PEA de 56,6%. Nota-se que o atraso de 5 horas da Taxa de Coque foi o que apresentou maior relevância para esta variável. Isto pode ser justificado em virtude da lentidão com que ocorre o processo e devido ao tempo médio de resposta desta variável, o que demanda um longo tempo para que todo o coque seja consumido. Esta ocorrência, bem como a ocorrência da Taxa de O<sub>2</sub> evidência a importância do consumo de coque para inserção de silício ao ferro-gusa. Já a predominância da ocorrência da Temperatura do ferro-gusa é justificada pela relação direta da temperatura para o processo.

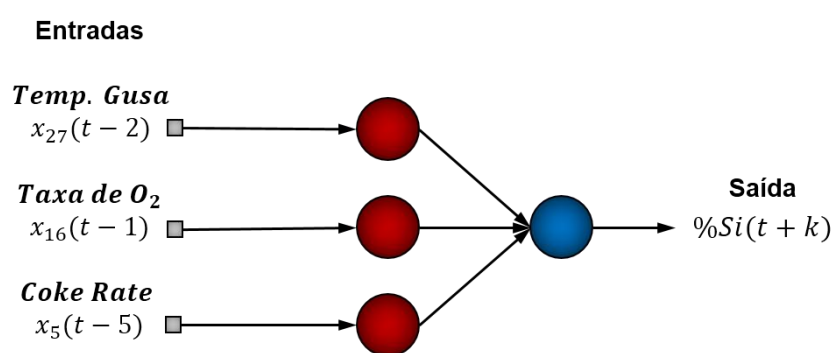
Ainda do Anexo B, observa-se que embora os resultados da implementação de número 8 tenham apresentado os melhores desempenhos (MSE de 0,0042, MAPE de 17,40% e PEA de 58,42%), trata-se de uma estrutura com um número de nós de entrada relativamente grande, comparada aos demais modelos resultantes. Comparando esta estrutura ao modelo obtido pela poda de número 9, nota-se que

com apenas três entradas e três neurônios ocultos foi possível obter um modelo capaz de prever com desempenho semelhante ao da oitava estrutura e, por esta razão, este será o modelo escolhido.

Em relação aos dados apresentados pelo Anexo C, deve-se destacar a forte ocorrência dos valores autorregressivos da série de silício frente às demais entradas para o modelo NARX. Conforme pode ser observado na Tabela 10, esta variável ocorreu cerca de 58% a mais que a segunda variável (Taxa de Coque). Neste sentido, é possível verificar a forte influência dos valores autorregressivos durante a modelagem. Outro detalhe importante foi observado na poda de número seis, na qual foi obtida uma rede com apenas os valores autorregressivos da série de silício, resultando assim, em um modelo NAR com 4 neurônios ocultos, MSE de 0,001, MAPE de 7,65% e PEA de 89,32%. No entanto, a estrutura que apresentou o melhor desempenho foi o modelo de número 9, com um MSE de 0,0011, MAPE de 6,41%, PEA de 93,82% e 6 neurônios ocultos e, portanto, a mesma será escolhida para aplicação das previsões.

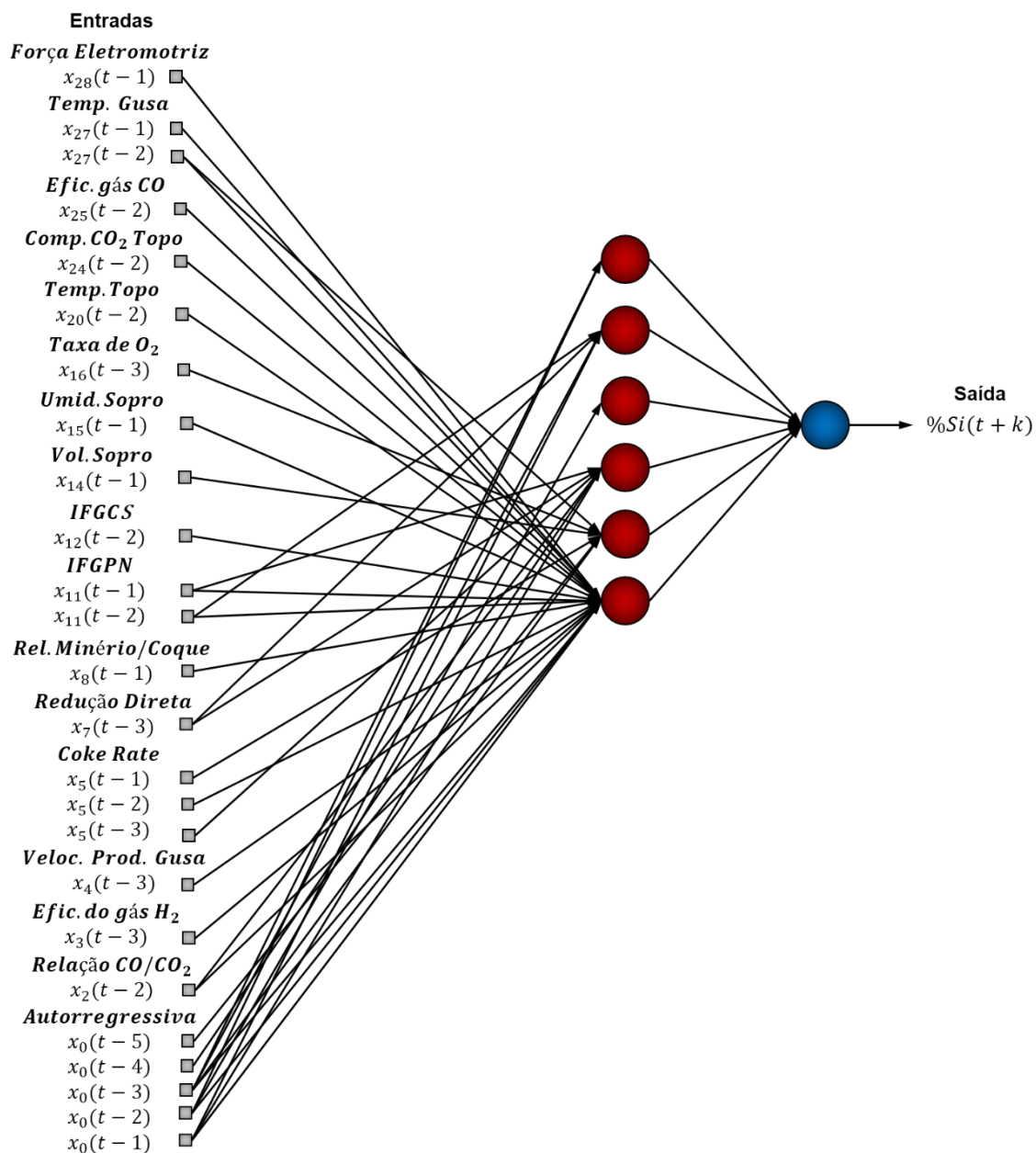
As Figuras 43 e 44 apresentam as redes neurais escolhidas após a aplicação das podas, respectivamente, nas redes NIO e NARX.

Figura 43 – Rede NIO resultante após aplicação do Algoritmo de Poda.



Fonte: O próprio autor.

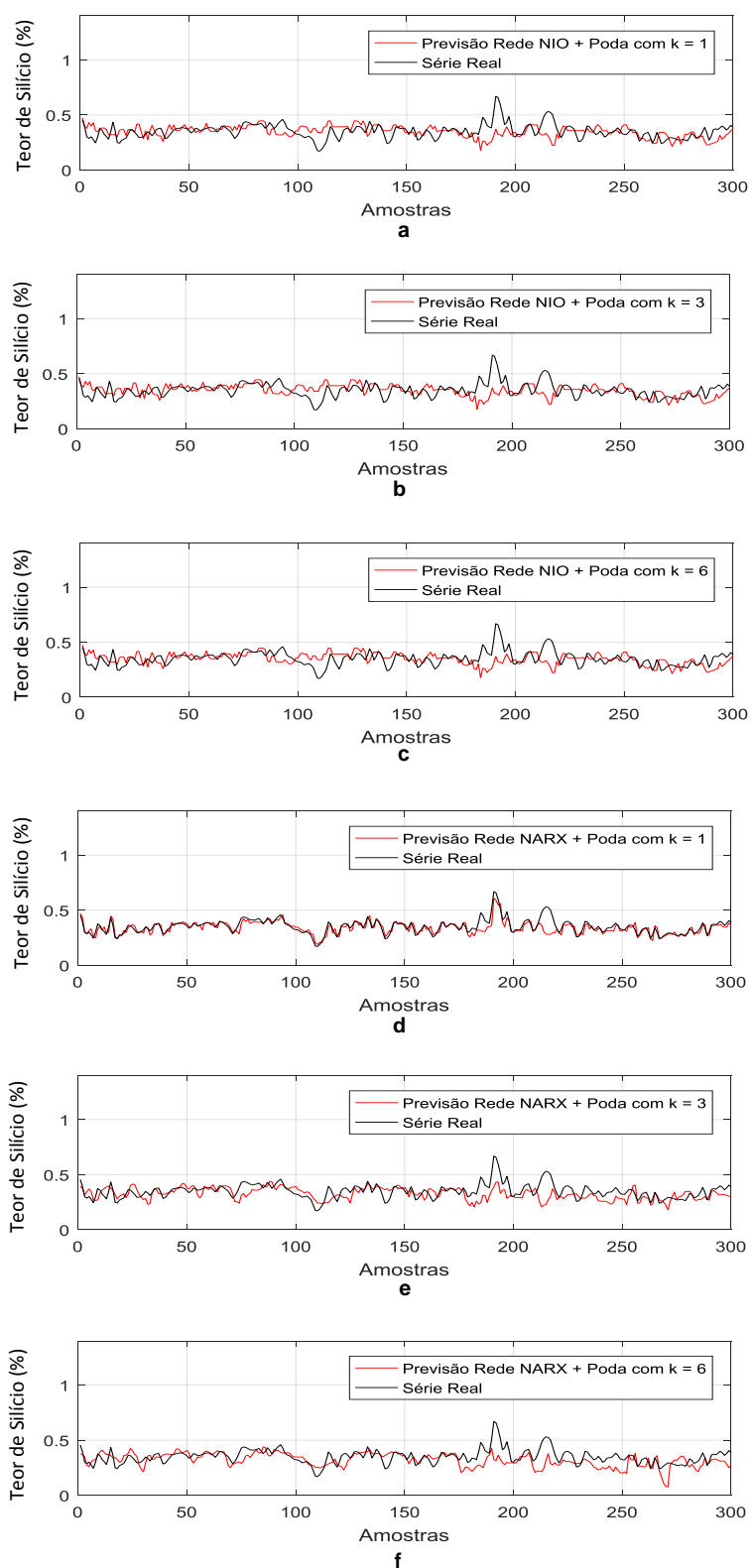
Figura 44 – Rede NARX resultante após aplicação do Algoritmo de Poda.



Fonte: O próprio autor.

As previsões do teor de silício obtidas por cada uma das topologias de rede com as entradas/atrasos selecionados após a aplicação do OBS, pode ser visualizado através da Figura 45.

Figura 45 – Previsão do Conteúdo de Silício do ferro-gusa aplicando Algoritmo de Poda: Rede NIO com horizonte de previsão de a) 1 hora; b) 3 horas; c) 6 horas; Rede NARX com horizonte de previsão de d) 1 hora; e) 3 horas; f) 6 horas.



Fonte: O próprio autor.

Além disto, na Tabela 11 são apresentados os respectivos desempenhos obtidos para as previsões de 1 a 6 passos à frente antes e depois da implementação do Algoritmo de Poda.

Tabela 11 – Comparativo dos critérios de desempenho das previsões antes e depois da aplicação do Algoritmo de Poda nas Redes NIO e NARX.

Modelo	Horizonte de Previsão	Antes do Algoritmo de Poda			Depois do Algoritmo de Poda		
		MSE	MAPE (%)	PEA (%)	MSE	MAPE (%)	PEA (%)
Rede NIO	1 hora	0,02178	39,88	28,651	0,00540	19,10	56,601
	2 horas	0,02163	39,25	28,410	0,00531	19,00	56,821
	3 horas	0,02115	38,85	27,323	0,00524	18,76	58,309
	4 horas	0,02054	38,28	30,606	0,00528	18,78	58,533
	5 horas	0,02029	38,05	30,932	0,00538	18,74	60,028
	6 horas	0,02048	38,25	28,147	0,00545	18,76	59,830
Rede NARX	1 hora	0,00418	17,58	55,617	0,00107	6,41	93,820
	2 horas	0,01517	34,41	30,942	0,00401	14,69	66,385
	3 horas	0,01342	32,15	33,802	0,00531	17,98	56,338
	4 horas	0,01165	29,55	36,530	0,00567	18,82	54,583
	5 horas	0,01237	30,42	34,745	0,00604	19,55	51,412
	6 horas	0,01187	29,04	36,633	0,00640	20,13	49,504

Fonte: O próprio autor.

Observa-se uma melhoria significativa no desempenho das redes, além de uma considerável diminuição no tamanho da camada oculta, por exemplo, a rede NARX que antes possuía 90 neurônios ocultos, com MSE de 0,01342, MAPE de 32,15% e PEA de 33,80% (horizonte de previsão de 3 horas), agora com 6 neurônios ocultos, foi obtido um PEA de aproximadamente 60% e uma redução do MAPE em cerca de 44%.

O mesmo efeito pode ser observado para o modelo empregando a Rede NIO. Considerando-se um horizonte de previsão de 3 horas, a rede que antes era composta por 95 neurônios ocultos e resultava em um MSE de 0,02115, MAPE de 38,85% e PEA de 27,323%, é capaz de prever com desempenho superior (um MSE de 0,00524, MAPE de 18,76% e PEA de 58,309%) utilizando, para isto, apenas 3 neurônios ocultos.

O Algoritmo de Poda também foi empregado ao modelo NAR, conforme apresenta a Tabela 12. Neste caso, considerando a expressiva redução do número de neurônios da camada de entrada e consequente redução da complexidade do modelo, para aplicação da poda, esta rede foi inicializada com 100 neurônios na camada oculta.

Tabela 12 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NAR.

<b>Poda</b>	<b>Neurônios Ocultos</b>	<b>MSE</b>	<b>MAPE (%)</b>	<b>PEA (%)</b>
1	38	0,000253	3,39	98,179
2	15	0,000293	3,50	97,759
3	45	0,000279	3,56	98,039
4	13	0,000238	3,28	98,739
5	17	0,00028	3,50	98,319
6	8	0,000286	3,44	98,039
7	25	0,000248	3,33	98,179
8	23	0,00028	3,44	97,759
9	35	0,000316	3,70	97,198
10	8	0,000259	3,28	97,759

Fonte: O próprio autor.

A partir da Tabela 12, nota-se que o tamanho da camada oculta pode ser reduzido de 100 neurônios para 13 neurônios ocultos (4ª implementação do Algoritmo de Poda), preservando-se o desempenho do modelo, conforme é verificado na Tabela 13. Embora nenhum atraso da camada de entrada tenha sido eliminado durante o processo de poda, observa-se uma considerável redução do número de neurônios ocultos da camadas oculta devido à eliminação de pesos pouco significativos, bem como ao ajuste destes valores.

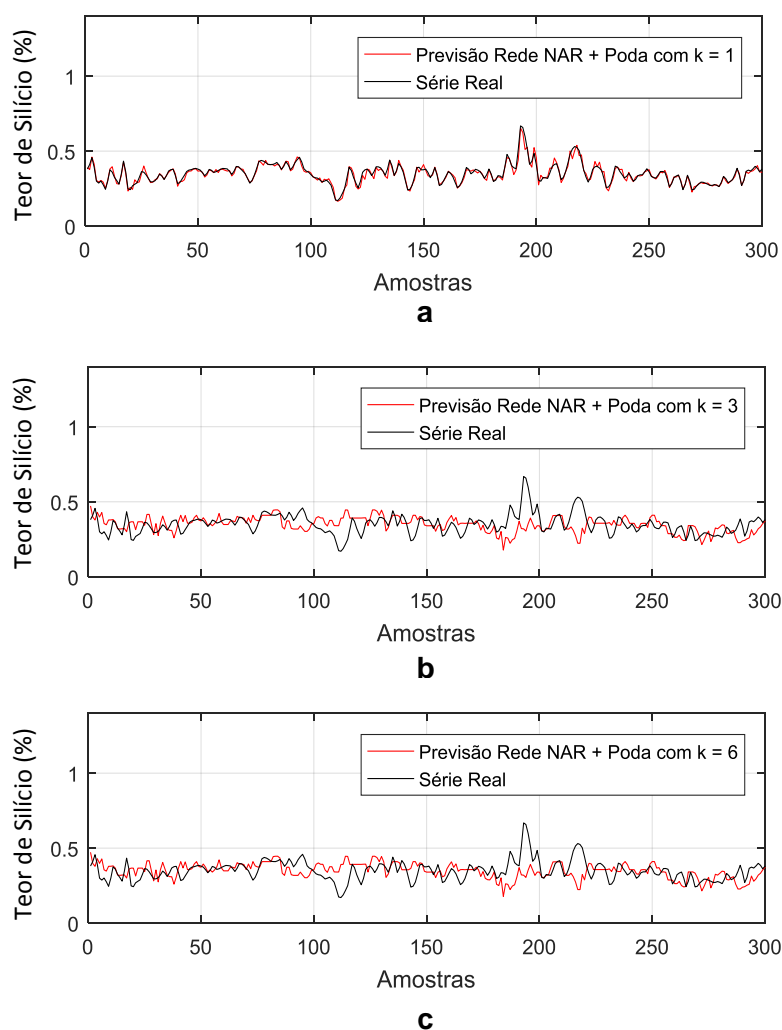
A Tabela 13 apresenta os desempenhos obtidos para as previsões de 1 a 6 passos à frente antes e depois da implementação do Algoritmo de Poda. Observa-se que mesmo com uma redução de 87 neurônios ocultos da rede, esta foi capaz de prever com desempenho semelhante ao modelo sem a poda. Já a Figura 46 ilustra as previsões de 1, 3 e 6 passos à frente obtidas a partir do modelo com Rede NAR após a aplicação da poda.

Tabela 13 – Comparativo dos critérios de desempenho das previsões antes e depois da aplicação do Algoritmo de Poda na Rede NAR.

Horizonte de Previsão	Antes do Algoritmo de Poda			Depois do Algoritmo de Poda		
	MSE	MAPE (%)	PEA (%)	MSE	MAPE (%)	PEA (%)
1 hora	0,000257	3,34	98,179	0,000238	3,28	98,739
2 horas	0,001221	7,93	87,237	0,001093	7,41	89,060
3 horas	0,003057	12,88	73,736	0,003365	12,29	76,264
4 horas	0,004406	15,07	66,948	0,005883	16,49	66,244
5 horas	0,009075	21,76	53,803	0,015318	26,26	52,535
6 horas	0,015177	29,50	40,480	0,02152	31,76	40,761

Fonte: O próprio autor.

Figura 46 – Previsão do Conteúdo de Silício do ferro-gusa aplicando Algoritmo de Poda: Rede NAR com horizonte de previsão de a) 1 hora; b) 3 horas; c) 6 horas.



Fonte: O próprio autor.

Em virtude da complexidade da série a ser modelada, visualmente a rede NAR não apresentou melhorias satisfatórias. Uma das possíveis causas para isso é a presença de atrasos devido aos termos autorregressivos que tendem a tornar as previsões mais inertes (WALLER; SAXEN, 2000).

Também foi observado que, conforme aumenta-se o horizonte de previsão, a precisão de cada rede diminuiu. O ruído e outros fatores podem distorcer a relação entre a série de teor de silício e suas variáveis de entrada. Além disto, com o aumento do horizonte de previsão, os modelos não são capazes de prever os valores extremos e tendem a fazer previsões próximas da média.

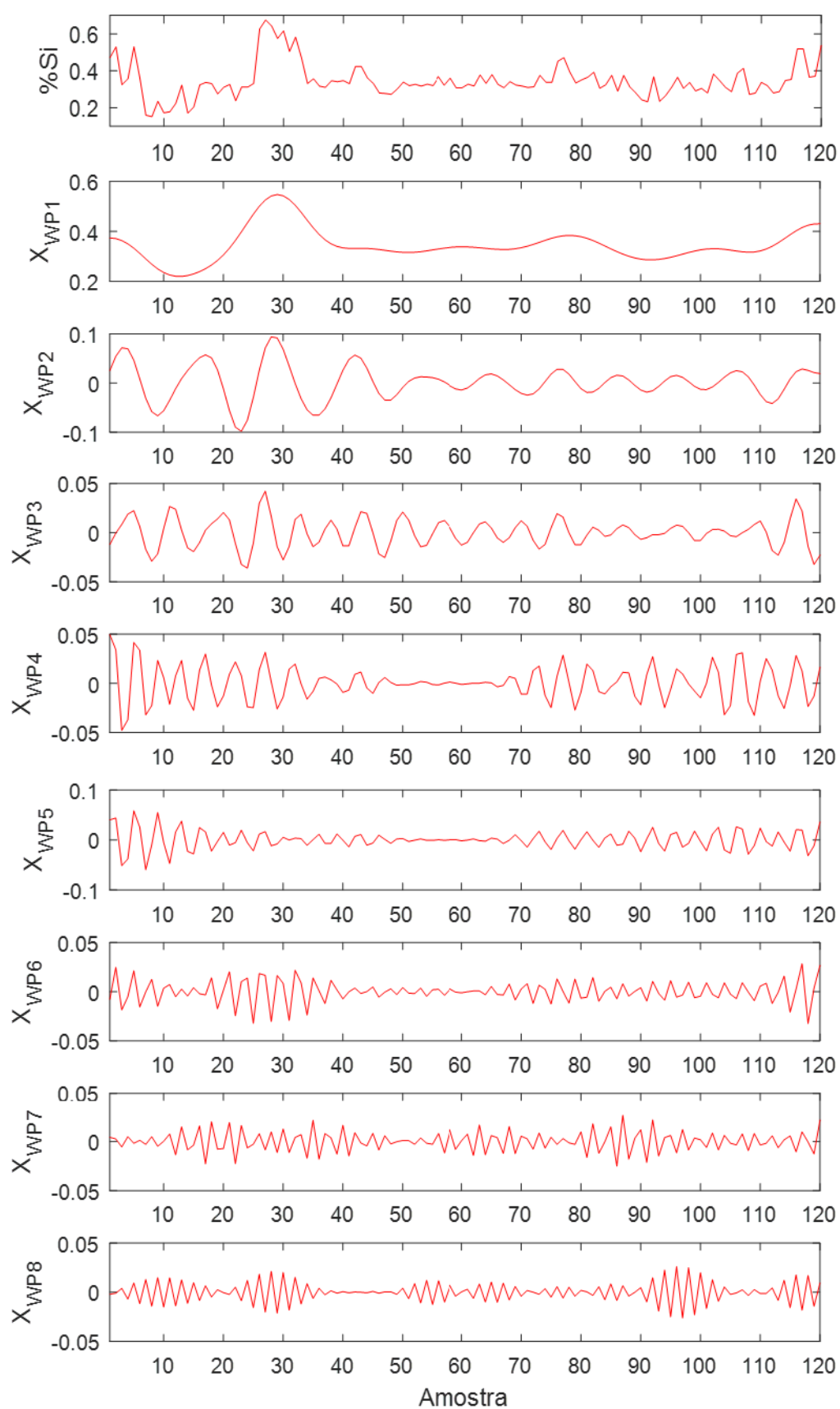
## 5.7 MODELAGEM E PREVISÃO DO TEOR DE SILÍCIO UTILIZANDO UM ALGORITMO HÍBRIDO DE MODWPT E NAR

Com o intuito de diminuir a complexidade da série de silício, propõe-se a aplicação da decomposição MODWPT. Primeiramente, o sinal original é decomposto em 3 níveis através da MODWPT, utilizando uma *wavelet* base do tipo Symlet de ordem 4. A Figura 47 apresenta a série original seguida dos resultados da decomposição de 120 amostras, apenas para fins de visualização das subséries. Observa-se que mesmo com a decomposição, todas as componentes apresentam contribuições significativas. A partir de então, são utilizadas oito configurações de redes NAR para fazer a previsão de cada uma das 8 subséries. Por fim, as previsões de cada subsérie são somadas, obtendo-se assim a previsão do sinal original. Em virtude da simplificação das subséries devido à decomposição, serão testados horizontes de previsão maiores, variando de 1 a 8 passos à frente.

A Figura 48 mostra a estrutura do algoritmo. Nesta figura, cada uma das subséries,  $X_{WP1}, X_{WP2}, \dots, X_{WP8}$ , são previstas respectivamente pelas redes,  $NAR_1, NAR_2, \dots, NAR_8$ , resultando nas subséries  $\hat{X}_{WP1}, \hat{X}_{WP2}, \dots, \hat{X}_{WP8}$ , que são somadas gerando a saída prevista,  $\hat{y}(t + k)$ , de acordo com o horizonte de previsão  $k$ .

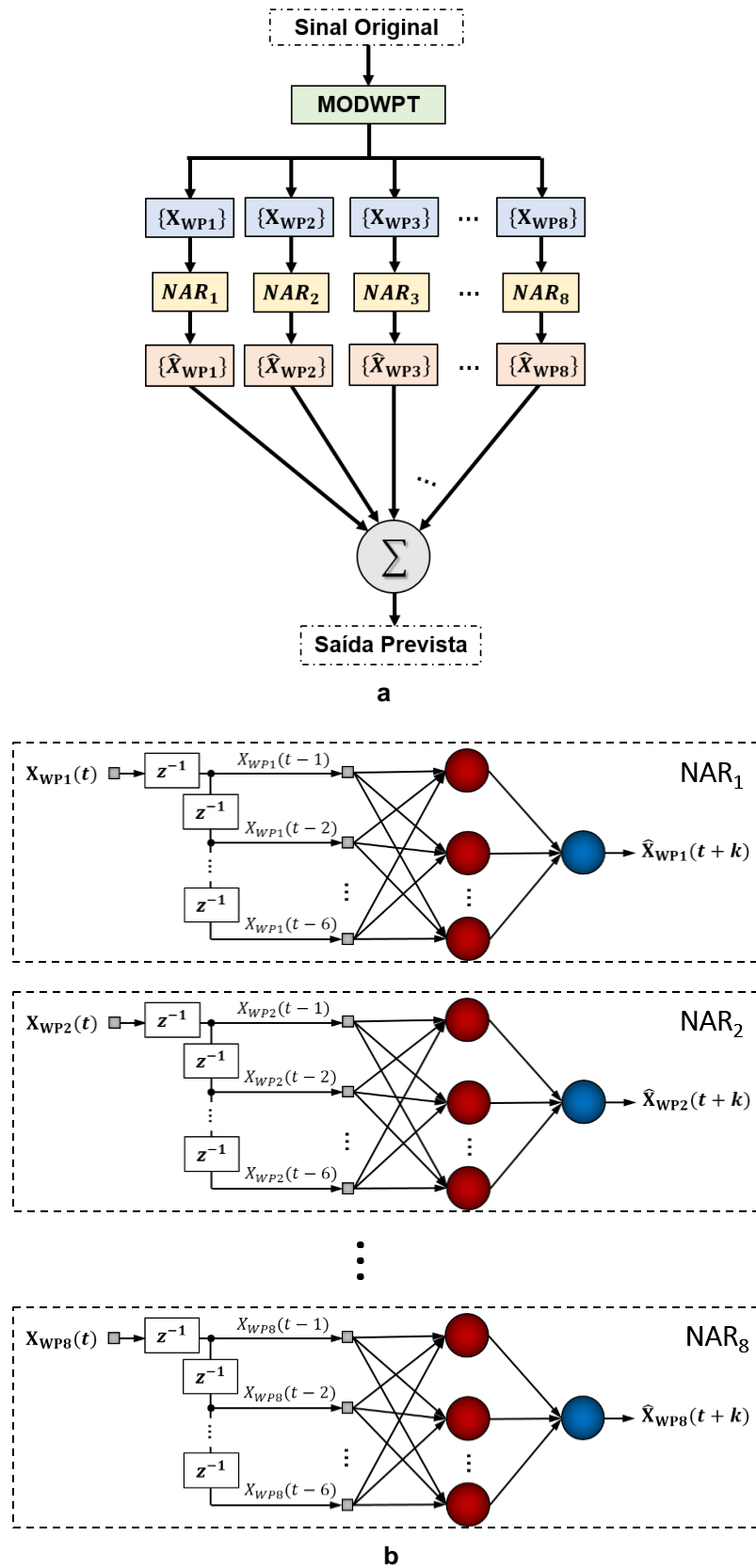


Figura 47 – A MODWPT: Série decomposta em 8 subséries.



Fonte: O próprio autor.

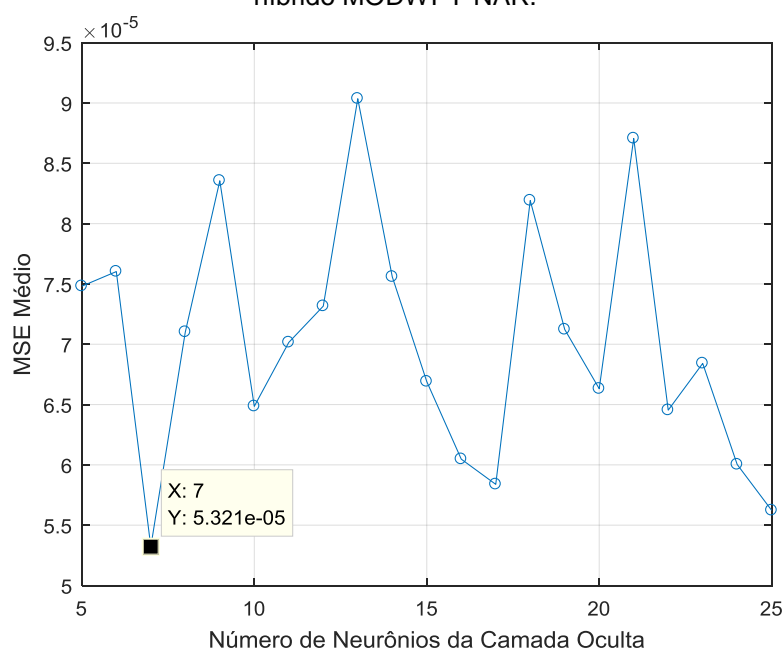
Figura 48 – Estrutura do algoritmo MODWPT-NAR.



Tal como para as demais redes apresentadas, sem a utilização do Algoritmo de Poda, para a escolha do número de neurônios da camada oculta que melhor se adequa a cada uma das redes NAR do modelo híbrido MODWPT-NAR, foi aplicado o teste da variação do número de neurônios ocultos. Este experimento foi dividido em duas partes e, para ambas, foram testadas redes com camadas ocultas variando de 1 a 25 neurônios.

A princípio, variou-se o número de neurônios ocultos igual para cada uma das redes NAR empregadas na previsão de cada uma das subséries geradas pela MODWPT. Estes resultados foram confrontados com o segundo teste, no qual foram testados diferentes tamanhos de camadas ocultas em cada uma das redes NAR separadamente. Uma vez que, cada série apresenta diferentes níveis de complexidade, esta segunda parte do teste é justificada pois, o tamanho da camada oculta de cada rede seria ajustado conforme a complexidade da subsérie empregada. A Figura 49 apresenta o MSE médio de treinamento resultante após variar-se igualmente o tamanho da camada oculta de todas as redes NAR, simultaneamente, conforme evidencia a primeira parte do experimento. De acordo com a mesma, o tamanho da camada oculta de todas as oito redes NAR, configuradas para previsão de cada uma das subséries, que resultou no menor MSE médio foi de 7 neurônios ocultos (MSE médio igual a  $5,321 \times 10^{-5}$ ).

Figura 49 – O MSE médio de treinamento obtido para cada tamanho de camada oculta do modelo híbrido MODWPT-NAR.



Fonte: O próprio autor.

Tabela 14 – Tamanho Da Camada Oculta da rede para a Previsão de cada uma das Subséries

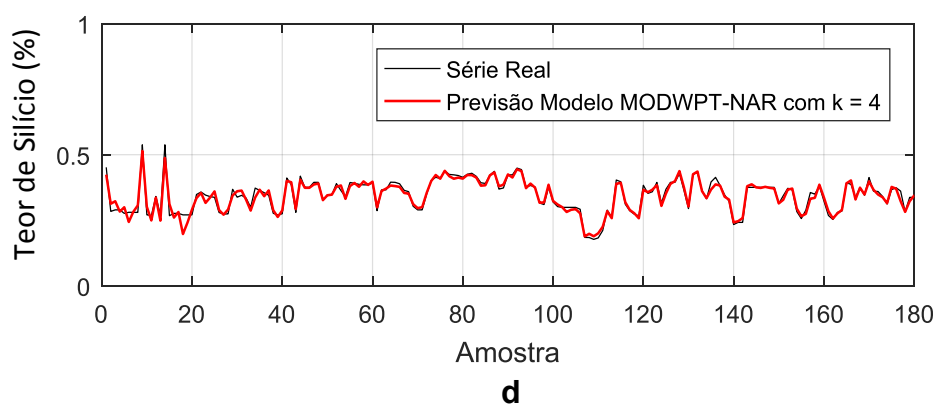
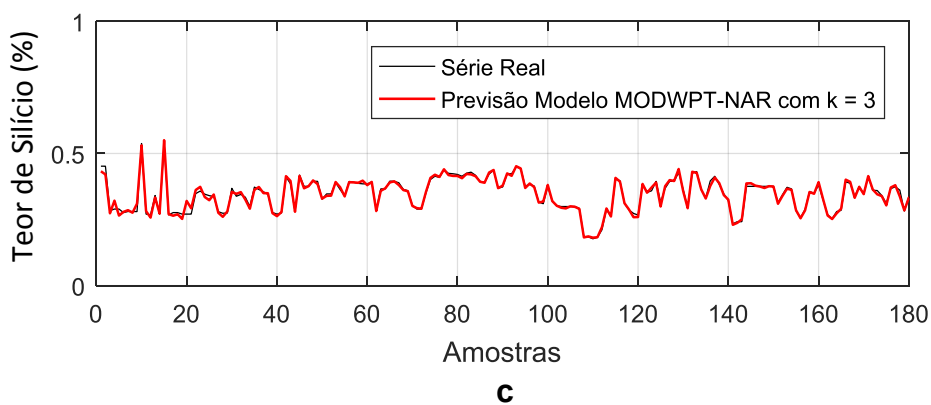
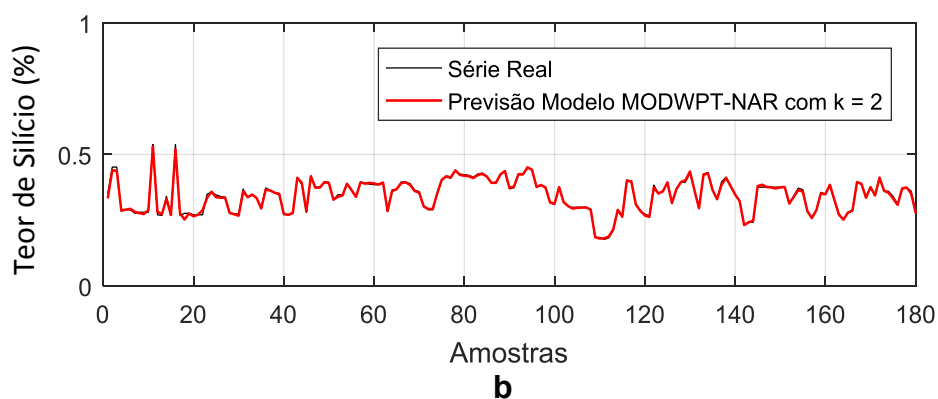
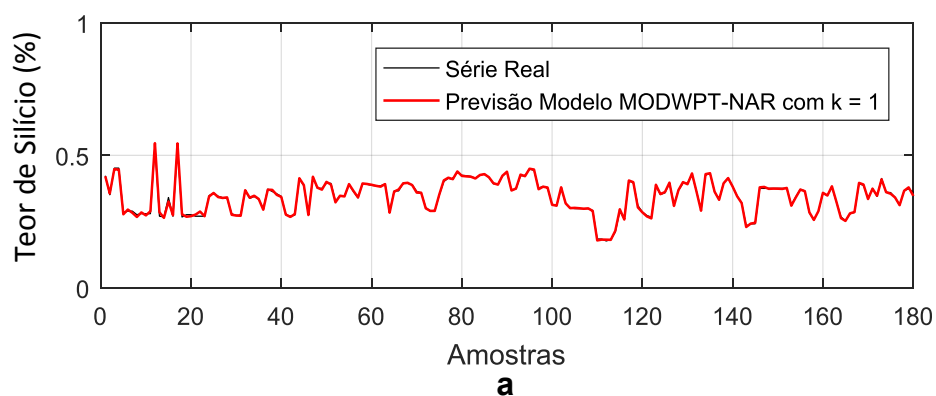
Subsérie	Rede Neural	Tamanho da Camada Oculta	MSE ( $\times 10^{-4}$ )
$X_{WP1}$	NAR <sub>1</sub>	27	0,0169
$X_{WP2}$	NAR <sub>2</sub>	8	0,2642
$X_{WP3}$	NAR <sub>3</sub>	7	0,1299
$X_{WP4}$	NAR <sub>4</sub>	19	0,0487
$X_{WP5}$	NAR <sub>5</sub>	21	0,0500
$X_{WP6}$	NAR <sub>6</sub>	25	0,0442
$X_{WP7}$	NAR <sub>7</sub>	23	0,0708
$X_{WP8}$	NAR <sub>8</sub>	14	0,0047
MSE de Treinamento Final (somatório)			0,6294

Fonte: O próprio autor.

Observa-se, no entanto, que os MSE apresentados pela Tabela 14 são referentes ao treinamento de cada uma das subséries separadamente. Avaliando-se, então, o MSE de treinamento final (somatório dos MSE resultantes de cada rede), foi obtido, para a segunda parte do teste, um valor de  $6,294 \times 10^{-5}$ . Confrontando este valor com aquele resultante da primeira parte do teste ( $5,321 \times 10^{-5}$ ), observa-se que redes de tamanhos relativamente menores foram capazes de prever com um MSE ligeiramente menor. Isto é justificado pois, durante a primeira parte do experimento, o tamanho das camadas ocultas de cada uma das redes foi escolhido a partir da avaliação do MSE de treinamento calculado a partir da série original e da série prevista, diferentemente do que ocorreu na segunda parte, em que o MSE de treinamento foi calculado a partir de cada subsérie e sua respectiva previsão. Assim, uma vez que, nesta segunda parte do teste, as redes não operam simultaneamente em conjunto, resultou-se em conexões redundantes.

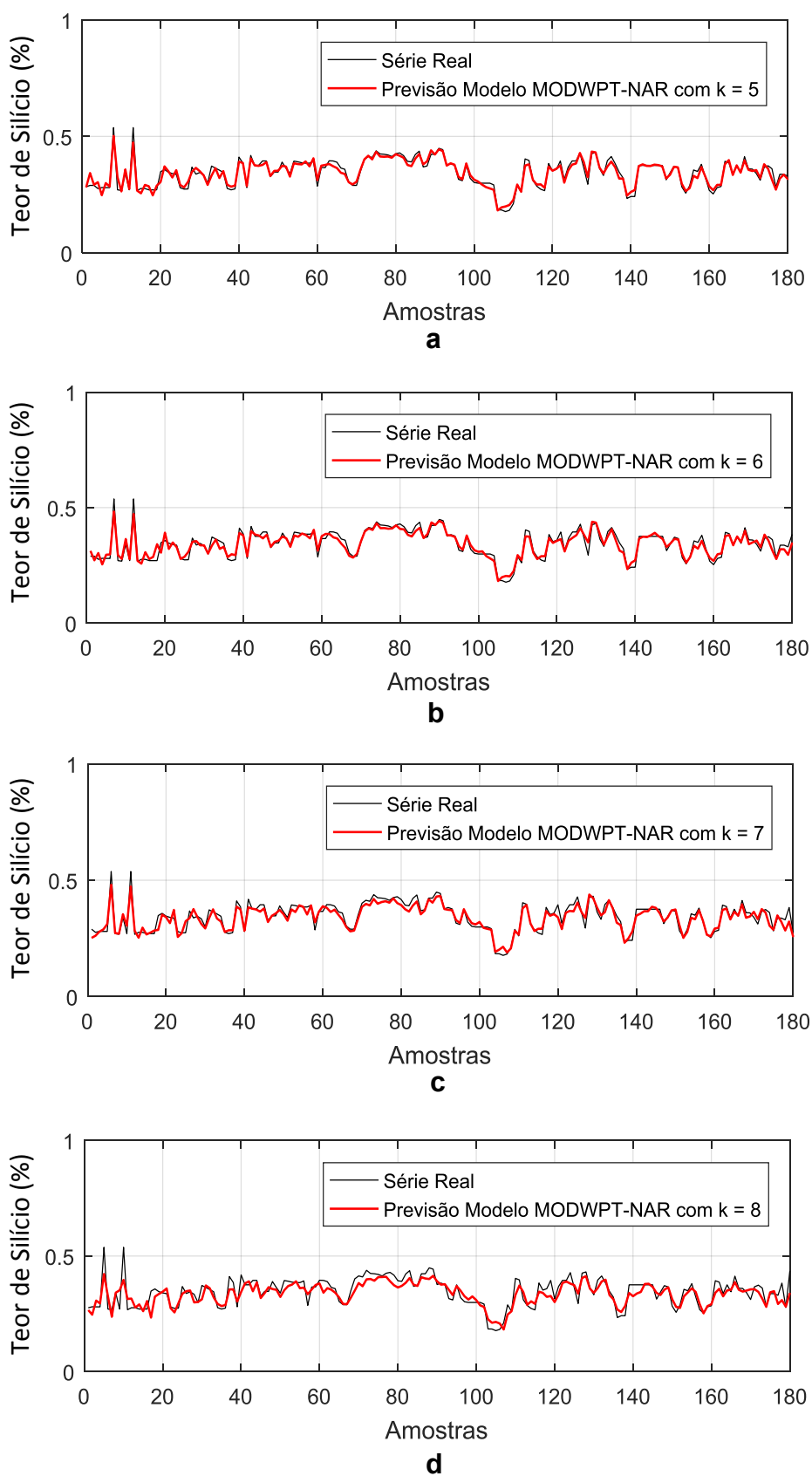
As Figuras 50 e 51 mostram os resultados obtidos na previsão de 1 a 8 passos à frente das 180 primeiras amostras de testes. Os desempenhos calculados por cada previsão são apresentados na Tabela 15, a qual também apresenta um comparativo com os mesmos resultados obtidos pela Rede NAR, resultante da poda, sem a aplicação da decomposição.

Figura 50 – Previsão do Conteúdo de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NAR com horizonte de previsão de a) 1 hora; b) 2 horas; c) 3 horas; d) 4 horas.



Fonte: O próprio autor.

Figura 51 – Previsão do Conteúdo de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NAR com horizonte de previsão de a) 5 horas; b) 6 horas; c) 7 horas; d) 8 horas.



Fonte: O próprio autor.

Tabela 15 – Comparativo entre Critérios de Desempenhos das Previsões de 1 a 8 passos à frente do Modelo MODWPT-NAR e do Modelo NAR.

<b>Modelo</b>	<b>Horizonte de Previsão</b>	<b>MSE</b>	<b>MAPE (%)</b>	<b>PEA (%)</b>
MODWPT-NAR	1 hora	0,0000083	0,68	100
	2 horas	0,0000168	0,99	100
	3 horas	0,0000429	1,58	100
	4 horas	0,0001376	2,82	99,72
	5 horas	0,0002210	3,54	99,02
	6 horas	0,0003019	4,21	98,18
	7 horas	0,0003380	4,54	97,90
	8 horas	0,0008323	6,57	93,70
NAR	1 hora	0,000238	3,28	98,73
	2 horas	0,001093	7,41	89,06
	3 horas	0,003365	12,29	76,26
	4 horas	0,005883	16,49	66,24
	5 horas	0,015318	26,26	52,53
	6 horas	0,02152	31,76	40,76
	7 horas	0,030023	37,42	37,71
	8 horas	0,038283	44,67	31,40

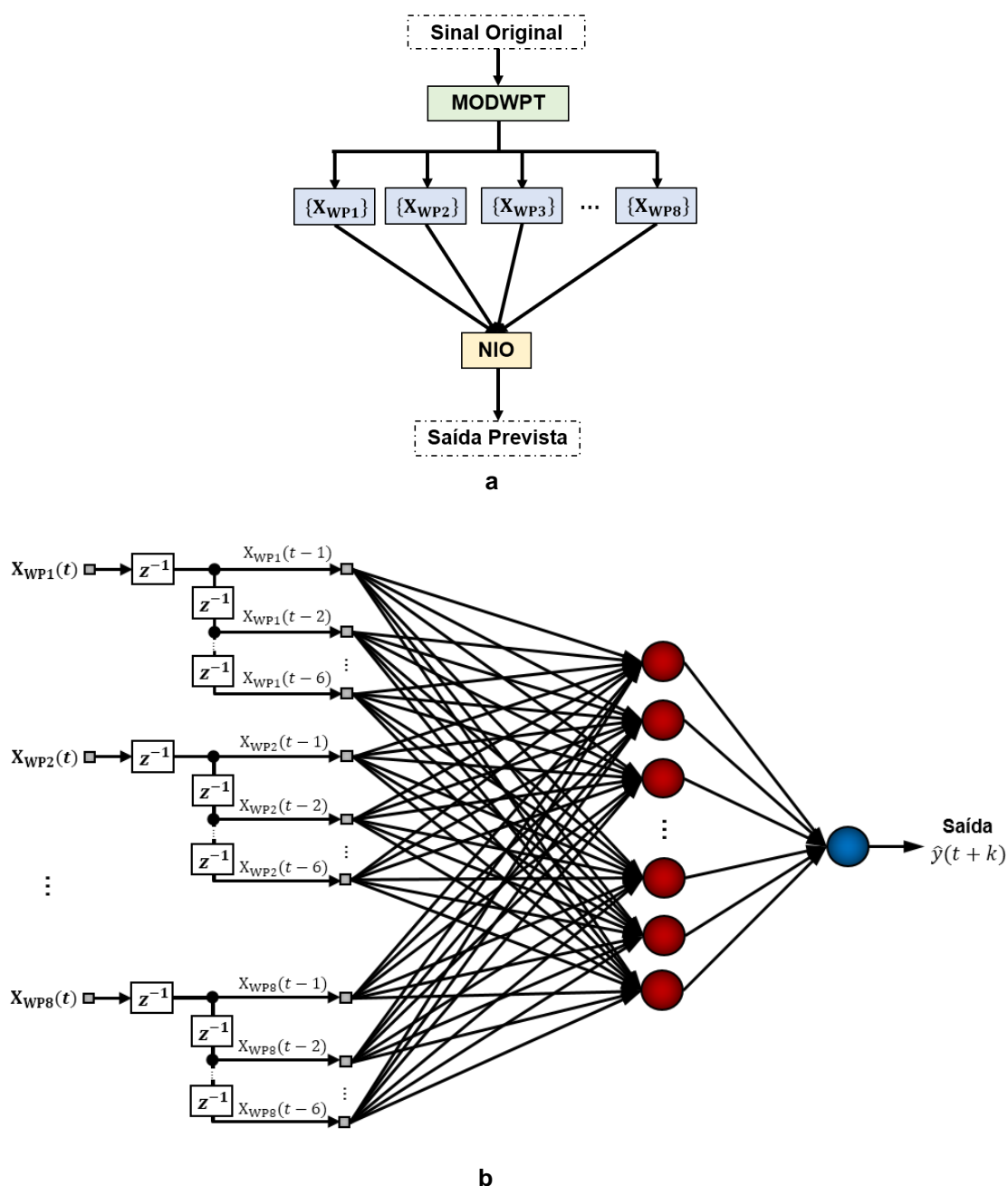
Fonte: O próprio autor.

Das Figuras 50 e 51 e da Tabela 15, pode ser observado que as previsões híbridas obtidas pelos modelos adotados neste estudo apresentaram desempenho superior na previsão do teor de silício no ferro-gusa. Tomando como exemplo os resultados da previsão de 3 passos à frente, o modelo combinado MODWPT-NAR obteve um MAPE de aproximadamente 1,58% contra 12,29% da previsão utilizando o modelo NAR. A razão para isto é que a MODWPT permite decompor o sinal não-estacionário em séries relativamente estáveis que reduzem a complexidade da previsão da RNA. Com relação ao desempenho dos modelos, conforme o horizonte de previsão aumenta, é possível observar, através da Tabela 15, que o desempenho das previsões do algoritmo híbrido não se degrada significativamente, diferentemente do que ocorre com o desempenho da previsão utilizando apenas a RNA. Além disto, avaliando o PEA, constata-se que 40,76% das amostras do modelo NAR estão dentro dessa faixa, contra 98,18% obtido pelo modelo combinado MODWPT-NAR, para uma previsão de 6 passos à frente.

## 5.8 MODELAGEM E PREVISÃO DO TEOR DE SILÍCIO UTILIZANDO UM ALGORITMO HÍBRIDO DE MODWPT E NIO

Outra abordagem proposta será a aplicação de uma única rede NIO de maneira a substituir as 8 redes do tipo NAR empregadas anteriormente. Deste modo, na camada de entrada da rede NIO serão aplicadas cada uma das subséries do sinal de silício com os seus respectivos regressores. Esta implementação será denominada MODWPT-NIO e está graficamente representada através da Figura 52.

Figura 52 – MODWPT-NIO: a) Estrutura do algoritmo; b) Implementação da RNA do tipo NIO.

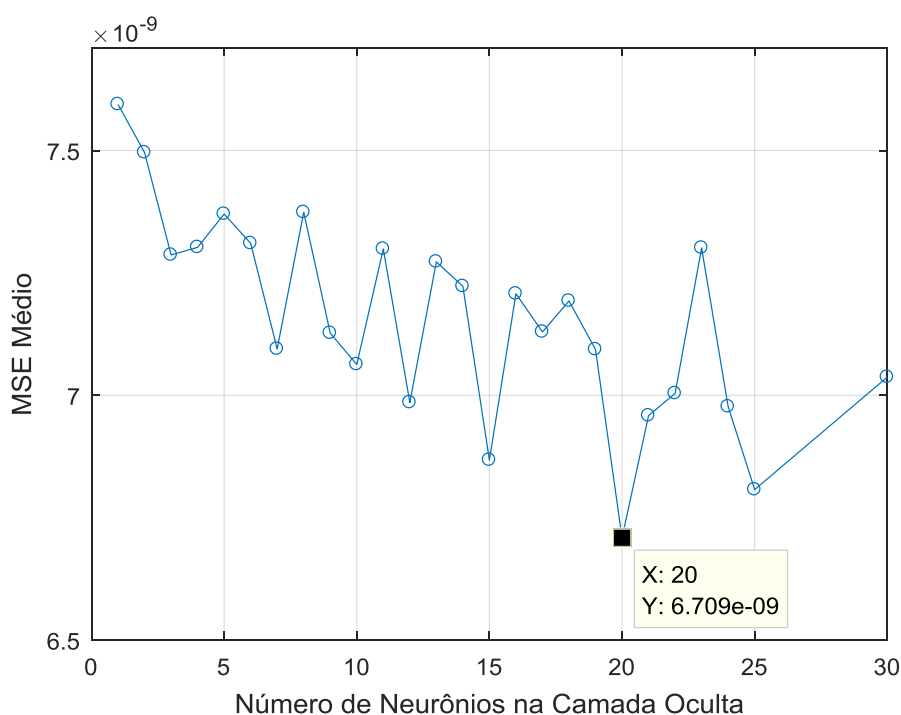


Fonte: O próprio autor.



Para a escolha do número de neurônios da camada oculta que melhor se adequa ao modelo híbrido MODWPT-NIO, foi utilizado o teste da variação do tamanho da camada oculta de 1 a 30 neurônios ocultos. A Figura 53 evidencia que o menor valor médio de MSE obtido durante o teste foi para uma rede de 20 neurônios ocultos, com o valor de  $6,709 \times 10^{-9}$ .

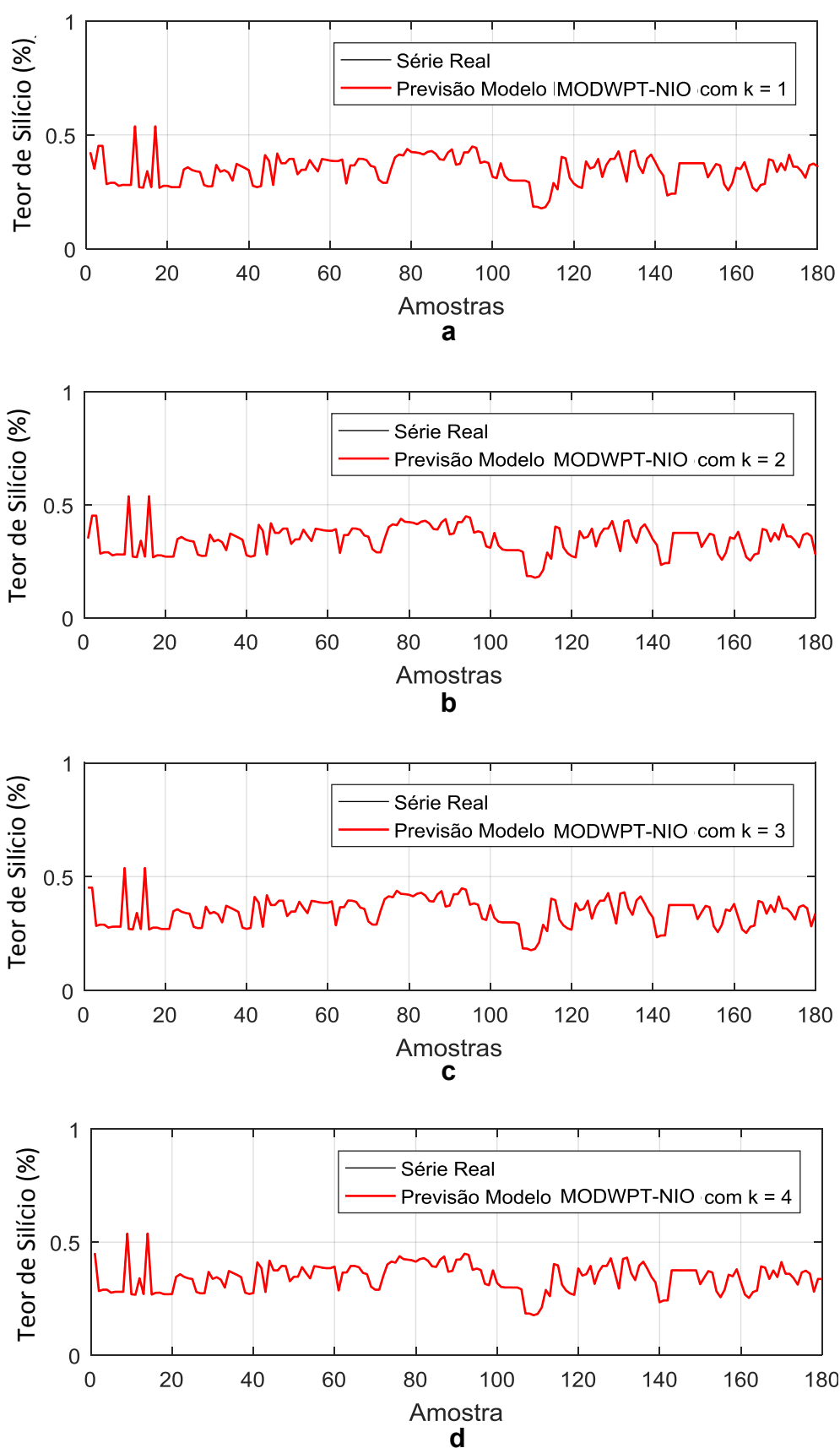
Figura 53 – O MSE médio de treinamento obtido para cada tamanho de camada oculta do modelo híbrido MODWPT-NIO.



Fonte: O próprio autor.

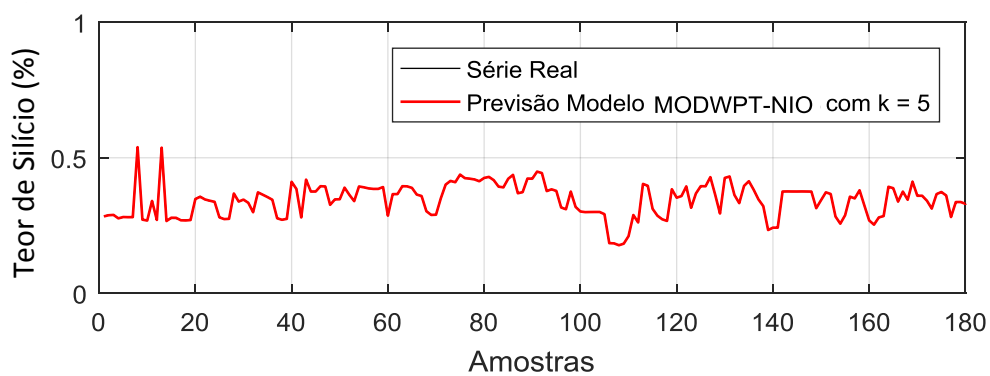
As Figuras 54 e 55 mostram os resultados obtidos na previsão de 1 a 8 passos à frente das 180 primeiras amostras de testes, enquanto que os desempenhos calculados por cada previsão são apresentados na Tabela 16, a qual também compara os resultados obtidos com aqueles obtidos pelo modelo MODWPT-NAR.

Figura 54 – Previsão do Conteúdo de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NIO com horizonte de previsão de a) 1 hora; b) 2 horas; c) 3 horas; d) 4 horas.

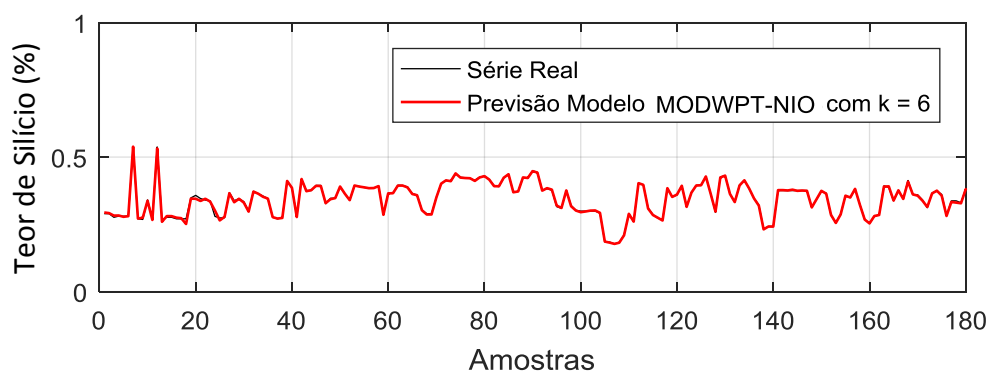


Fonte: O próprio autor.

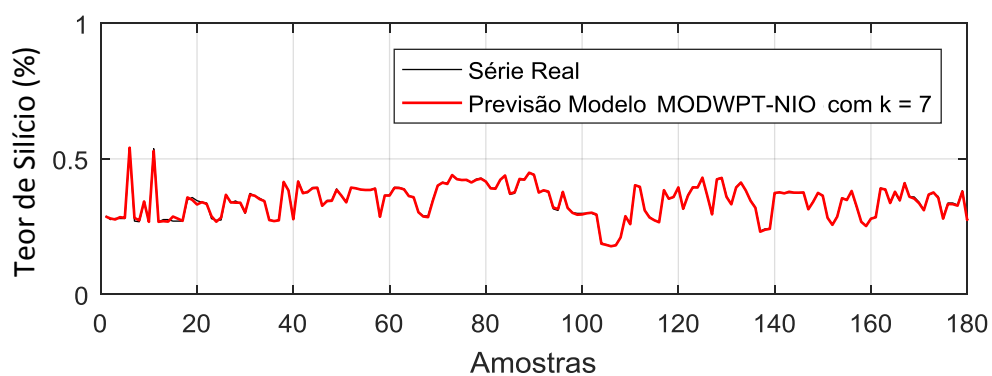
Figura 55 – Previsão do Conteúdo de Silício do ferro-gusa empregando um modelo híbrido MODWPT-NIO com horizonte de previsão de a) 5 horas; b) 6 horas; c) 7 horas; d) 8 horas.



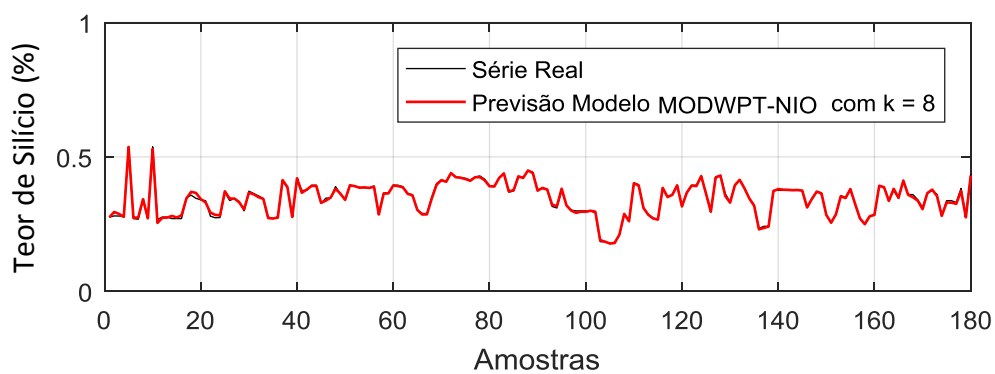
**a**



**b**



**c**



**d**

Fonte: O próprio autor.

Tabela 16 – Comparativo entre Critérios de Desempenhos das Previsões de 1 a 8 passos à frente dos Modelos MODWPT-NAR e MODWPT-NIO.

Modelo	Horizonte de Previsão	MSE	MAPE (%)	PEA (%)
MODWPT-NIO	1 hora	$1,88 \times 10^{-10}$	0,0030	100
	2 horas	$1,69 \times 10^{-9}$	0,0085	100
	3 horas	$1,45 \times 10^{-8}$	0,0267	100
	4 horas	$1,04 \times 10^{-7}$	0,0764	100
	5 horas	$2,36 \times 10^{-7}$	0,1117	100
	6 horas	0,0000067	0,4947	100
	7 horas	0,0000074	0,6159	100
	8 horas	0,0000173	0,8626	99,86
MODWPT-NAR	1 hora	0,0000083	0,6881	100
	2 horas	0,0000168	0,9921	100
	3 horas	0,0000429	1,5857	100
	4 horas	0,0001376	2,8231	99,72
	5 horas	0,0002210	3,5466	99,02
	6 horas	0,0003019	4,2157	98,18
	7 horas	0,0003380	4,5463	97,90
	8 horas	0,0008323	6,5745	93,70

Fonte: O próprio autor.

Assim como no modelo MODWPT-NAR, o desempenho das previsões do modelo MODWPT-NIO não se degradou significativamente com o aumento do horizonte de previsão. No entanto, é possível observar a superioridade da previsão do modelo MODWPT-NIO frente ao modelo MODWPT-NAR. Avaliando-se o PEA, para uma previsão de 6 passos à frente, constata-se que 100% das amostras do modelo MODWPT-NIO estão dentro dessa faixa, correspondendo à cerca de 2% a mais que o modelo MODWPT-NAR.

Acredita-se que o modelo MODWPT-NIO foi capaz de prever melhor o teor de silício do ferro-gusa para horizontes de previsão mais altos, fazendo uso de uma rede relativamente pequena, visto que esta rede leva em consideração a série original como referência na fase de treinamento.

## 5.9 COMPARATIVO ENTRE OS RESULTADOS OBTIDOS COM MODELOS ENCONTRADOS NA LITERATURA

A Tabela 17 apresenta um comparativo entre os desempenhos obtidos pelo modelo MODWPT-NIO para uma previsão de um, três, seis e oito passos à frente com outros métodos sugeridos na literatura. É possível notar que o modelo proposto neste trabalho apresentou os melhores índices de desempenho, mesmo sem a utilização de entradas exógenas. No entanto, é importante frisar que os dados utilizados nas modelagens propostas pelas literaturas não foram os mesmos usados neste estudo.

Tabela 17 – Comparativo entre os resultados gerados pelas diferentes metodologias encontradas na literatura.

Modelo	Período de Amostragem (min)	Horizonte de Previsão	Entradas Exógenas	MSE	MAPE (%)	PEA (%)
MODWPT-NIO	60	1	Não	$1,88 \times 10^{-10}$	0,0030	100
		3	Não	$1,45 \times 10^{-8}$	0,0267	100
		6	Não	0,0000067	0,4947	100
		8	Não	0,0000173	0,8626	99,86
Saxén e Pettersson (2007)	60	2	Sim	0,074	-	-
Chen (2001)	48	3	Sim	0,0062	14,3	-
Chen, Wang e Han (2010)	144	1	Sim	-	-	95,7
Nurkkala, Pettersson e Saxén (2011)	60	3	Sim	0,0845	-	-
Wang <i>et al.</i> (2015)	-	1	Não	2,3628	-	-

Fonte: O próprio autor.

Considerando os resultados dos modelos que utilizaram apenas entradas exógenas para a previsão do teor de silício, foi obtido um MSE de 0,0062 e um MAPE de 14,3% (CHEN, 2001). Outros estudos apresentaram um MSE de aproximadamente 0,074 (SAXÉN; PETTERSSON, 2007) e 0,0845 (NURKKALA; PETTERSSON; SAXÉN,

2011). É importante frisar, no entanto, que no trabalho proposto por Chen (2001) não foi informado o intervalo de normalização aplicado aos dados, comprometendo-se assim o comparativo entre os MAPE.

No trabalho de Chen, Wang e Han (2010), que além das entradas exógenas, também utilizou o valor atual do percentual de silício no ferro-gusa, foi obtido um PEA de 95,7%, entretanto, este estudo não apresentou os respectivos valores de MSE e MAPE.

Em relação ao trabalho de Wang *et al.* (2015), que adotou o EMD para decompor a série de silício em diferentes séries temporais de frequência, foi obtido um MSE de 2,3628. Deve-se mencionar que nos estudos apresentados por Saxén e Peterson (2007), Chen (2001), Nurkkala, Pettersson e Saxén (2011) e Wang *et al.* (2015) não foi mensurado o valor do PEA.

Ainda de acordo com as informações apresentadas pela Tabela 17, é possível confirmar a superioridade dos resultados gerados pelo modelo MODWPT-NIO, que mesmo prevendo com um horizonte de 8 horas à frente foi capaz de obter um MSE muito menor do que aqueles obtidos pelos demais modelos propostos pela literatura, utilizando horizontes de previsão menores.

Outro fator que deve ser levado em consideração é que devido a utilização, muitas vezes, de somente variáveis exógenas e por conta das particularidades inerentes aos processos das siderurgias nos artigos citados, foram realizadas previsões de até 3 horas à frente, perfazendo um horizonte de previsão mais simples. No entanto, por conta da simplificação da série temporal através da MODWPT, o modelo proposto por este trabalho foi capaz de prever o teor de silício com até 8 horas de antecedência com alta qualidade de previsão e resultados mais promissores que os obtidos com os modelos citados nos artigos com horizontes de previsão de até 3 horas.

## 6 CONCLUSÃO

Inicialmente, foram propostos modelos empregando três topologias de redes neurais (NARX, NIO e NAR) e, com o propósito de atenuar o nível de ruído inerente às medições das variáveis visando a diminuição dos erros de previsão, foram avaliados três tipos de filtros. De uma maneira geral, as redes NIO, NARX e NAR que apresentaram as melhores previsões, foram aquelas que utilizaram os dados de entradas filtrados não-linearmente através de uma *wavelet*-mãe do tipo Symlet de ordem 4.

Contudo, devido à complexidade do processo de alto-forno, um grande número de variáveis envolvidas no processo podem afetar no teor de silício. Neste contexto, visando a seleção destas variáveis, foi aplicado um Algoritmo de Poda do tipo OBS após o treinamento das redes neurais. Ao comparar os modelos NARX e NIO resultantes da poda qualitativamente quanto às suas capacidades de previsão de até 6 horas à frente, foi constatada uma melhoria dos critérios de desempenho, além da redução do número de neurônios da camada oculta e do número de entradas de cada rede.

De fato, a regularidade do funcionamento do alto-forno permite que a RNA reconheça o sistema a partir de variáveis do processo. No entanto, as ações do operador nestas variáveis, visando corrigir momentos de instabilidade do alto-forno e/ou manter o teor de silício dentro da faixa aceitável, podem provocar variações indesejadas nesta série temporal, aumentando o seu comportamento não-estacionário e, conseqüentemente, diminuindo a capacidade de aprendizagem da rede neural. Por estes motivos, as previsões não apresentaram desempenhos muito significativos em horizontes de previsão maiores que 3 horas.

Visando contornar os efeitos da não-estacionariedade da série de silício na modelagem da rede neural, foi proposta a aplicação da decomposição MODWPT em 3 níveis, resultando em 8 subséries com características mais estacionárias em relação à série de silício. Estas subséries foram aplicadas em duas configurações híbridas distintas (MODWPT-NAR e MODWPT-NIO) e as previsões de cada uma delas, de 1 a 8 horas à frente, foram comparadas.

A princípio, notou-se a superioridade das previsões híbridas frente àquelas obtidas pelos demais modelos anteriormente propostos ao longo deste trabalho. Notou-se também que o modelo MODWPT-NIO foi o que apresentou os resultados mais promissores, mesmo para horizontes de previsão mais altos, visto que este modelo

leva em consideração somente a série original durante a fase de treinamento, enquanto que o modelo MODWPT-NAR utiliza cada uma das subséries como referência no treinamento.

Ao confrontar os resultados obtidos pelo modelo MODWPT-NIO com encontrados na literatura, foi possível verificar a superioridade dos resultados apresentados nesta dissertação, mesmo trabalhando com horizontes de previsões maiores.

O modelo MODWPT-NIO demonstrou ser uma ferramenta promissora para a previsão do teor de silício do ferro-gusa e, portanto, sugere-se, para trabalhos futuros, uma abordagem mais específica acerca desta técnica, bem como a possibilidade de sua implementação na planta.



## REFERÊNCIAS

- ADAMOWSKI, J. F. River flow forecasting using wavelet and cross-wavelet transform models. **Hydrological processes**, v. 22, n. 25, p. 4877-4891, 2008.
- AGUIRRE, L. A. **Introdução à identificação de sistemas: técnicas lineares e não-lineares: teoria e aplicação**. 4. ed. Belo Horizonte: Ed. UFMG, 2015.
- ALVES, D. K. *et al.* Real-Time Power Measurement Using the Maximal Overlap Discrete Wavelet-Packet transform. **IEEE Transactions on Industrial Electronics**, v. 64, n. 4, p. 3177-3187, 2017.
- ARAÚJO, L. A. **Manual de Siderurgia**. São Paulo: Editora Arte & Ciência, 1997.
- ARCELORMITTAL Tubarão. Ficha Técnica Alto Forno 03. **Fluxo de produção**. Disponível em <<http://tubarao.arcelormittal.com/quem-somos/usina/fluxo-producao/popfluxo.htm>>. Acesso em: 28 out. 2017.
- ARRIVABENE, L. F.; MACHADO, M. L. P.; SOBRINHO, V. F. M. **Siderurgia para não siderurgistas**. Vitória: Centro Federal de Educação Tecnológica do Espírito Santo, 2003.
- BAG, S. K. ANN based prediction of blast furnace parameters, **Journal – The Institution of Engineers**, Malaysia, v. 68, n. 1, p. 37-42, mar. 2007.
- BARROS, M. **Processos Estocásticos**. Rio de Janeiro: Papel Virtual Editora, 2004.
- BEALE, M. H.; HAGAN, M. T.; DEMUTH, H. B.; **Neural Network Toolbox™: User's Guide**. Natick: Mathworks, Inc., 2017.
- BISHOP, C. M. **Neural networks for pattern recognition**. New York: Oxford University Press, 1995.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**, 3rd ed. San Francisco: Holden-Day, 1976.
- BRAGA, A. P.; LUDEMIR, T.B.; CARVALHO, A. C. P. L. F. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC Editora, 2014.
- BUENO, R. L. S. **Econometria de Séries Temporais**. 2. ed. São Paulo: Cengage Learning, 2011.
- BULSARI, A.; SAXEN, H.; SAXEN, B. Time-series prediction of silicon in pig iron using neural networks. In: International Conference on Engineering Applications of Neural Networks (EANN '92), 1992.
- CAMPOS, V. F. **Tecnologia de fabricação do aço líquido**. 2. ed. Belo Horizonte: UFMG, 1983.

CHEN W.; WANG, B. X.; HAN, H. L. Prediction and control for silicon content in pig iron of blast furnace by integrating artificial neural network with genetic algorithm. **Ironmaking & Steelmaking**, v. 37, n. 6, p. 458–463, ago. 2010.

CHEN, J. A predictive system for blast furnaces by integrating a neural network with qualitative analysis. **Engineering Applications of Artificial Intelligence**, v. 14, n. 1, p. 77-85, fev. 2001.

CHUNG, Y. H.; CHOU, R. J.; CHANG-CHIANG, G. M.; CHEN, G. L. An autoregressive model for the system dynamics of blast furnace. **Journal China Inst. Chem. Eng.**, v. 17, n. 2, p. 107–111, mar. 1986.

CORANI, G.; GUARISO, G. An application of pruning in the design of neural networks for real time flood forecasting. **Neural Computing & Applications**, v. 14, n. 1, p. 66-77, 2005.

CORDUAS, M. Nonlinearity tests in time series analysis. **Journal of the Italian Statistical Society**, v. 3, n. 3, p. 291-313, 1994.

DAUBECHIES, I. Orthonormal bases of compactly supported wavelets. **Communications on pure and applied mathematics**, v. 41, n. 7, p. 909-996, 1988.

DAUBECHIES, I. **Ten lectures on wavelets**. Philadelphia: Society for industrial and applied mathematics, 1992.

DE BOOR, C. **A Practical Guide to Splines**. New York: Springer-Verlag, 1978.

DHOND, A.; GUPTA, A.; VADHAVKAR, S. Data mining techniques for optimizing inventories for electronic commerce. In: ACM SIGKDD international conference on Knowledge discovery and data mining, 6th, 2000, Boston. **Anais...** Boston: ACM, 2000, p. 480-486.

DHOND, Anjali. **Application of neural network techniques for modeling of blast furnace parameters**. 2000. 96 f. Dissertação (Mestrado em Engenharia Elétrica e Ciência da Computação) - Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Massachusetts, 2000.

DINIZ, A. P. M.; GOMES, F. S. V.; SALLES, J. L. F.; COCO, K. F. **Uso de modelos NAR e ARIMA para previsão da série de teor de silício no ferro-gusa**. Trabalho apresentado no XIII Simpósio Brasileiro de Automação Inteligente, Porto Alegre, 2017. Disponível em: <[https://www.ufrgs.br/sbai17/papers/paper\\_346.pdf](https://www.ufrgs.br/sbai17/papers/paper_346.pdf)>. Acesso em: 25 jan. 2018.

DORFFNER, G. Neural networks for time series processing. **Neural Network World**, v. 6, n. 4, p. 447–468, 1996.

DUNNE, S.; GHOSH, B. Weather Adaptive Traffic Prediction Using Neurowavelet Models. **IEEE Transactions on Intelligent Transportation Systems**, v. 14, n. 1, p. 370-379, 2013.

FIGUEIRA, R. M.; CASTRO, L. F. A.; TAVARES, R. P. **Controle do teor de silício no alto-forno**. Belo Horizonte: Fundação Christiano Ottoni, 1989.

FUN, M. H.; HAGAN, M. T. Levenberg-marquardt training for modular networks. In: IEEE International Conference on Neural Networks, 1996, Washington. **Anais...** Washington: IEEE Press., 1996, p. 468-473.

GABOR, D. Theory of communication. **Journal of the Institution of Electrical Engineers**. v. 93, n. 26, p. 429-457, 1946.

GE, A. X. **A Neural Network Approach to the Modeling of Blast Furnace**. 1999. 70 f. Tese (Doutorado em Engenharia Elétrica e Ciências da Computação) - Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Massachusetts, 1999.

GEERDES, M.; TOXOPEUS, H.; VAN DER VLIET, C. **Modern Blast Furnace Ironmaking**: an introduction. 2nd ed. Amsterdam: IOS Press. 2009.

GIBBONS, J. D.; CHAKRABORTI, S. **Nonparametric Statistical Inference**. 4th ed. New York: Marcel Dekker, Inc., 2003.

GOMES, Flávio da Silva Vitorino. **Modelos de séries temporais para previsão de nível de líquido em cadinho de altos-fornos**. 2016. 109 f. Tese (Doutorado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Espírito Santo, Vitória, 2016.

GRAPS, A. An introduction to wavelets. **IEEE computational science and engineering**, v. 2, n. 2, p. 50-61, 1995.

GROSSMANN, A.; MORLET, J. **Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape**. SIAM Journal on Mathematical Analysis, v. 15, n. 4, p. 723-736, 1984.

GUSTAVSSON, Joel. **Reactions in the Lower Part of the Blast Furnace with Focus on Silicon**. 2004. 68 f. Tese (Doutorado em Ciências dos Materiais e Engenharia) - Division of Applied Process Metallurgy, Royal Institute of Technology, Stockholm, 2004.

HAAR, A. Zur Theorie der Orthogonalen Funktionensysteme. **Mathematische Annalen**, v. 69, n. 3, p. 331-371, 1910.

HAMILTON, J. **Time Series Analysis**. 1st. ed. New Jersey: Princeton University Press, 1994.

HASHEMI, A. A. N.; CLUM, J. A.; WU, S. M. Blast furnace modeling and control by the DDS method. In: Ironmaking Conference, 36th, 1977, Pittsburgh. **Anais...** Pittsburgh: AIME, 1977, p. 283–297.

HASSIBI, B.; STORK, D. G.; WOLFF, G. Optimal Brain Surgeon and general network pruning. In: IEEE International Conference on Neural Networks, 1993, San Francisco. **Anais...** San Francisco: IEEE Press., 1993, p. 293-299.

HAYKIN, S. **Redes Neurais: princípios e prática**. 2. ed, Porto Alegre: Bookman, 2001.

HAYKIN, S.; VENN, B. V. **Sinais e Sistemas**. Porto Alegre: Bookman, 2001.

HERTZ, J., KROGH, A.; PALMER, R. G. **Introduction to the theory of neural computation**. Redwood City: Addison-Wesley, 1991.

HORNIK, K. Approximation capabilities of multilayer feedforward networks. **Neural networks**, v. 4, n. 2, p. 251-257, 1991.

HU, Y. H.; HWANG, J. **Handbook of neural network signal processing**. Boca Raton: CRC Press, 2002.

ISMAIL, B.; YUNUS, M. Time Series Forecasting Using Undecimated Wavelets, Neural Networks and Genetic Algorithm. **International Journal of Electronics and Computer Science Engineering**, v. 1, n. 3, p.1404-1415, 2012.

KIRCHGÄSSNER, G.; WOLTERS, J. **Introduction to modern time series analysis**. Berlin: Springer Science & Business Media, 2007.

KOLMOGOROV, A. Sulla determinazione empirica di una legge di distribuzione. **Giornale dell'Istituto Italiano degli Attuari**, v. 4, p. 83–91, 1933.

KOSKO, B. **Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence**. New Jersey: Prentice hall, 1992.

KRUSKAL, W.; WALLIS, W. Use of ranks in one-criterion variance analysis. **Journal of the American Statistical Association**, v. 47, n. 260, p. 583–621, 1952.

KWIATKOWSKI, D.; PHILLIPS, P. C. B.; SCHMIDT, P.; SHIN, Y. T. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? **Journal of econometrics**, v. 54, n. 1-3, p. 159-178, 1992.

LECUN, Y., DENKER, J. S., SOLLA, S. A., HOWARD, R. E.; JACKEL, L. D. Optimal brain damage. In: TOURETZKY, D. (Ed.). **Advances in Neural Information Processing Systems**. Denver: Morgan Kaufmann, 1990, p. 598-605.

LEONTARITIS, I. J.; BILLINGS, S. A. Input-output parametric models for non-linear systems Part I: deterministic non-linear systems. **International Journal of Control**, v. 41, n. 2, p. 303-328, 1985.

LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. **Quarterly of applied mathematics**, v. 2, n. 2, p. 164-168, 1944.

LEWIS, C. D. **Demand Forecasting and Inventory Control**. New York: Wiley, 1997.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. **Journal of the American statistical Association**, v. 62, n. 318, p. 399-402, 1967.

LIU, H.; TIAN, H-Q.; PAN, D-F.; LI, Y-F. Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. **Applied Energy**, v. 107, p. 191–208, 2013.

MALLAT, S. **A wavelet tour of signal processing**. Burlington: Academic Press, 2008.

MALLAT, S. G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 7, p. 674-693, 1989.

MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. **Journal of the society for Industrial and Applied Mathematics**, v. 11, n. 2, p. 431-441, 1963.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115-133, 1943.

MISITI, M.; MISITI, Y.; OPPENHEIM, G.; POGGI, J. M. **Wavelets and their Applications**. Boston: John Wiley & Sons, 2013.

MISITI, M.; MISITI, Y.; OPPENHEIM, G.; POGGI, J-M. **Wavelet Toolbox User's Guide**. Natick: The MathWorks, Inc., 2017.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2. ed. São Paulo: Editora Blücher, 2006.

MOURÃO, M. B. *et al.* **Introdução à Siderurgia**. São Paulo: Associação Brasileira de Metalurgia, Materiais e Mineração, 2011.

MURTAGH, F.; STARCK, J-L.; RENAUD, O. On neuro-wavelet modeling. **Decision Support Systems**, v. 37, n. 4, p. 475-484, 2004.

NARENDRA, K. S.; PARTHASARATHY, K. Identification and control of dynamical systems using neural networks. **IEEE Transactions on neural networks**, v. 1, n. 1, p. 4-27, 1990.

NORGAARD, M. **Neural Network Based System Identification Toolbox**. Version 2. Denmark: Department of Automation, Technical University of Denmark, 2000.

NORGAARD, M.; RAVN, O.; HANSEN, L.; POULSEN, N. The nnsysid toolbox - a matlab toolbox for system identification with neural networks. In: IEEE Symposium on Computer-Aided Control System Design, 1996, Dearborn. **Anais...** Dearborn: IEEE Press., 1996, p. 374-379.

NORGAARD, M.; RAVN, O.; POULSEN, N. K.; HANSEN, L. K. **Neural networks for modelling and control of dynamic systems**. London: Springer 2000.

NURKKALA, A.; PETTERSSON, F.; SAXÉN, H. Non linear modeling method applied to prediction of hot metal silicon in the ironmaking blast furnace. **Ind. Eng. Chem. Res.**, v. 50, n. 15, p. 9236–9248, ago. 2011.

OLIVEIRA, H. M. **Análise de Fourier e Wavelets**: sinais estacionários e não estacionários. Recife: Ed. Universitária da UFPE, 2007a.

OLIVEIRA, H. M. **Análise de Sinais para Engenheiros**: Uma Abordagem via Wavelets. Recife: Brasport. 2007b.

OPPENHEIM, A. V., SCHAFER, R. W.; BUCK, J. R. **Discrete-Time Signal Processing**. Upper Saddle River: Prentice-Hall, 1999.

OPPENHEIM, A. V.; WILSKY, A. S. **Sinais e Sistemas**. 2. ed. São Paulo: Pearson Prentice Hall, 2010.

PANDIT, S. M.; CLUM, J. A.; WU, S. M. Modeling, prediction and control of blast furnace operation from observed data by multivariate time series. In: Ironmaking Conference, 34th, 1975, Toronto. **Anais...** Toronto: AIME, 1975, p. 403–416.

PANKRATZ, A. **Forecasting with Dynamic Regression Models**. Wiley-Interscience, 1991.

PERCIVAL D. B., GUTTORP P. Long-memory processes, the Allen Variance and Wavelets. In: FOUFOULA-GEORGIOU, E.; KUMAR P. (Eds.). **Wavelets in Geophysics**. San Diego: Academic Press, 1994, p. 325-344.

PERCIVAL, D. B.; MOFJELD, H. O. Analysis of subtidal coastal sea level fluctuations using wavelets. **Journal of the American Statistical Association**, v. 92, n. 439, p. 868-880, 1997.

PERCIVAL, D. B.; WALDEN, A. T. **Wavelet Methods for Time Series Analysis**. Cambridge: Cambridge University Press, 2000.

PESQUET, J.-C.; KRIM, H.; CARFANTAN, H. Time-invariant orthonormal wavelet representations. **IEEE transactions on signal processing**, v. 44, n. 8, p. 1964-1970, 1996.

POLIKAR, R. **The wavelet tutorial**. Ames: IOWA State University, 1996.

PRINCIPE, J. C.; EULIANO, N. R.; LEFEBVRE, W. C. **Neural and adaptive systems: fundamentals through simulations**. New York: John Wiley & Sons, Inc., 2000.

RADHAKRISHNAN, V. R.; MOHAMED, A. R. Neural networks for the identification and control of blast furnace hot metal quality. **Journal of process control**, v. 10, n. 6, p. 509-524, 2000.

REED, R. Pruning algorithms-a survey. **IEEE transactions on Neural Networks**, v. 4, n. 5, p. 740-747, 1993.

RENAUD, O.; MURTAGH, F.; STARCK, J-L. Wavelet-based forecasting of short and long memory time series. **Cahiers du Département d'économétrie**. Faculté des sciences économiques et sociales, Université de Genève. mai. 2002. Disponível em: <[http://www.unige.ch/ses/dsec/repec/files/2002\\_04.pdf](http://www.unige.ch/ses/dsec/repec/files/2002_04.pdf)>. Acesso: 22 Nov. 2017.

RIOUL, O.; VETTERLI, M. Wavelets and signal processing. **IEEE signal processing magazine**, v. 8, n. 4, p. 14-38, 1991.

RIZZO, E. M. da S. **Processo de fabricação de ferro gusa em alto-forno**. São Paulo: Associação Brasileira de Metalurgia, Materiais e Mineração, 2009.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533-538, 1986.

SARANGAPANI, J. **Neural network control of nonlinear discrete-time systems**. Boca Raton: CRC Press, 2006.

SAXÉN, H. Short-term prediction of silicon content in pig iron. **Canadian metallurgical quarterly**, v. 33, n. 4, p. 319-326, 1994.

SAXEN, H.; GAO, C.; GAO, Z. Data-Driven Time Discrete Models for Dynamic Prediction of the Hot Metal Silicon Content in the Blast Furnace – A Review. **IEEE Transactions on Industrial Informatics**, v. 9, n. 4, p. 2213-2225, nov. 2013.

SAXÉN, H.; PETTERSSON, F. Nonlinear prediction of the hot metal silicon content in the blast furnace, **ISIJ Int.**, v. 47, n. 12, p. 1732–1737, Dez. 2007.

SIETSMA, J; DOW, R. J. F. Neural net pruning-why and how. In: IEEE International Conference on Neural Networks, 1988. San Diego. **Anais...** San Diego: IEEE, 1988. p. 325-333.

SINGH, H.; SRIDHAR, N. V.; DEO, B. Artificial neural nets for prediction of silicon content of blast furnace hot metal. **Steel Research International**, v. 67, n. 12, p. 521-527, 1996.

SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. **The annals of mathematical statistics**, v. 19, n. 2, p. 279-281, 1948.

STEARNS, P. N. **The industrial Revolution in world history**. 3rd ed. Boulder: Westview Press, 2007.

THE NNSYSID Toolbox - For use with Matlab. 2003. Disponível em: <<http://www.iau.dtu.dk/research/control/nnsysid.html>>. Acesso em 04 abr. 2017.

TUNÇKAYA, Y.; KÖKLÜKAYA, E. Comparative performance evaluation of blast furnace flame temperature prediction using artificial intelligence and statistical methods. **Turkish Journal of Electrical Engineering & Computer Sciences**, v. 24, n. 3, p. 1163-1175, mar. 2016.

WALDEN, A. T.; CRISTAN, A. C. The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. **Proceedings of the Royal Society of London A**, v. 454, n. 1976, p. 2243–2266, 1998.

WALDEN, A.T. Wavelet Analysis of Discrete Time Series. **European Congress of Mathematics**, v. 202, p. 627-641, 2001.

WALLACE, J. P. *et al.* The Blast Furnace facility and equipment. In: WAKELIN, D. H. (Ed.). **The Making, Shaping and Treating of Steel**. 11th ed. Pittsburgh: The AISE Steel Foundation, 1999, p. 643-698.

WALLER, M.; SAXEN, H. On the development of predictive models with applications to a metallurgical process. **Industrial & engineering chemistry research**, v. 39, n. 4, p. 982-988, abr. 2000.

WANG, H.; YANG, G.; PAN, C.; GONG, Q. Prediction of Hot Metal Silicon Content in Blast Furnace based on EMD and DNN. In: Chinese Control Conference, 34th, 2015, Hangzhou. **Anais...** Hangzhou: IEEE, 2015, p. 8214-8218.

WANG, Y.; LIU, X. Prediction of Silicon Content in Hot Metal Based on SVM and Mutual Information for Feature Selection. **Journal of Information & Computational Science**, v. 8, n. 16, p. 4275-4283, 2011.

WEI, S.; ZUO, D.; SONG, J. Improving prediction accuracy of river discharge time series using a Wavelet-NAR artificial neural network. **Journal of Hydroinformatics**, v. 14, n. 4, p. 974-991, 2012.

WICKERHAUSER, M. V. **Adapted wavelet analysis from theory to software**. Wellesley: A.K. Peters, Ltd., 1994.

WIDROW, B.; HOFF, M. E. Adaptive switching circuits. In: ANDERSON, J. A.; ROSENFELD, E. (Eds.). **Neurocomputing: foundations of research**. Cambridge: J. MIT Press, 1988, p. 123-134.

WILAMOWSKI, B. M; IRWIN, J. D. **The Industrial Electronics Handbook: Intelligent systems**. Boca Raton: CRC Press, 2011.

YAN-JIONG, Z.; LIANG-TU, Y.; MIN, Z. Adopt three criterions to choose MISO prediction model of blast furnace process. In: IFAC Automation and Mining, Mineral and Metal Processing, 5th, 1986, Tokyo. **Anais...** Tokyo: IFAC, 1986, p. 241–246.

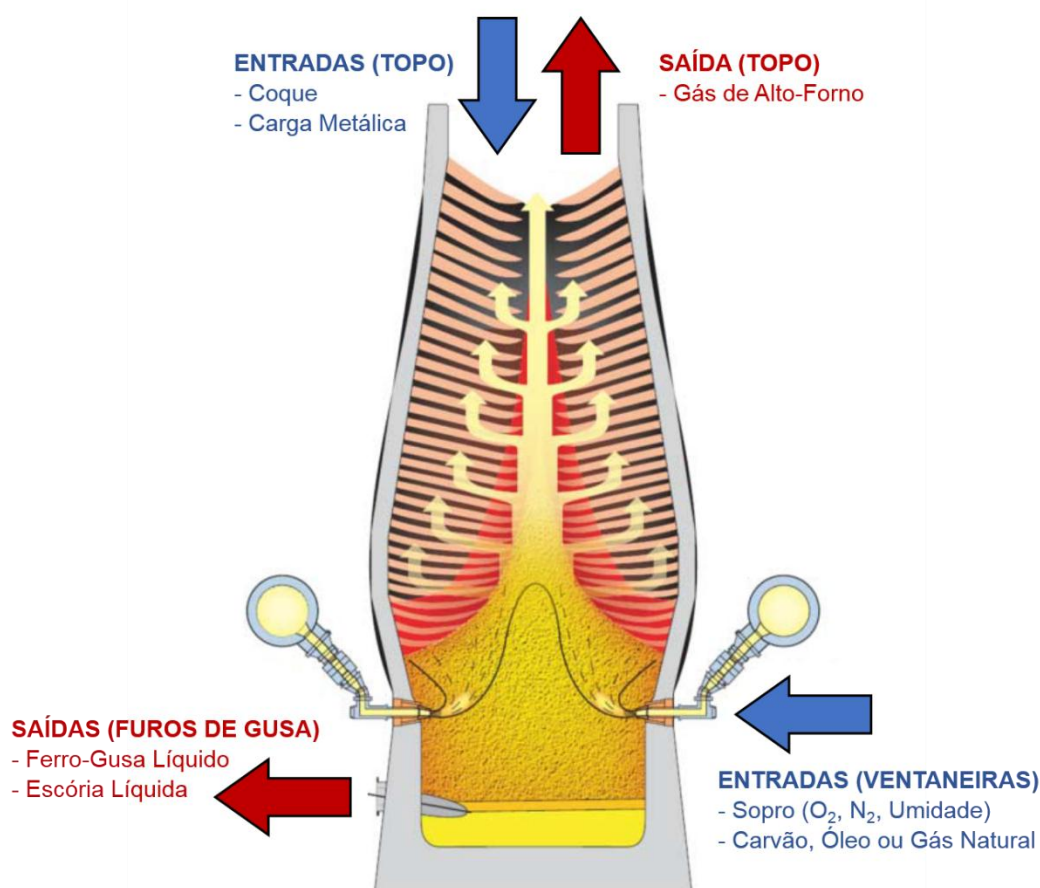
ZURADA, J. M. **Introduction to artificial neural systems**. St. Paul: West, 1992.



## APÊNDICE A: PRODUÇÃO DE FERRO-GUSA EM ALTO-FORNO

O alto-forno é um grande reator termo-químico empregado na produção de ferro-gusa, principal matéria-prima para a produção do aço. Conforme apresentado pela Figura 56, o funcionamento do alto-forno ocorre pelo princípio da contracorrente. Pela parte superior do alto-forno, são carregadas as matérias-primas ferrosas (minério de ferro, pelota e sinter), combustível (coque ou carvão vegetal) e fundentes (calcário, dolomita e quartzo). Na parte inferior, efetua-se o sopro de ar quente pelas ventaneiras, podendo haver também injeções auxiliares de combustível (óleo, alcatrão, carvão pulverizado ou gás natural), com o intuito de gerar calor ao processo através da combustão do carbono (ARAUJO, 1997; MOURÃO *et al.*, 2011).

Figura 56 – Representação do Alto-Forno com suas entradas e saídas.



Fonte: Adaptado de Geerdes, Toxopeus e Van Der Vliet (2009, p. 4).

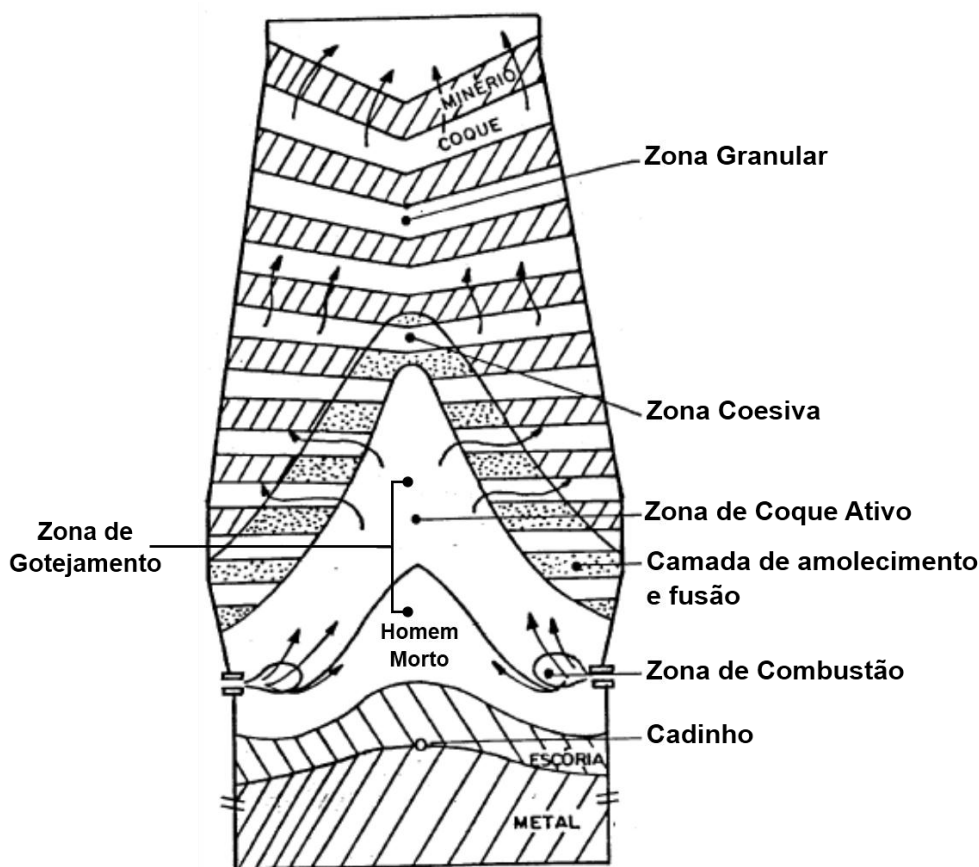
O objetivo deste processo é produzir uma liga, denominada de ferro-gusa, a uma temperatura de aproximadamente  $1500^{\circ}C$ , no estado líquido, composta de ferro (entre

92% a 95%), carbono (entre 3,5% a 5%) e outros elementos de liga como o silício (entre 0,3% e 0,7%), manganês (entre 0,2% e 0,4%), fósforo (entre 0,06% e 0,13%) e enxofre (tipicamente abaixo de 0,03%). Além disto, são também gerados, como subprodutos, a escória e o gás de alto-forno (GEERDES; TOXOPEUS; VAN DER VLIET, 2009; MOURÃO *et al.*, 2011).

### A.1 ASPECTOS INTERNOS DO ALTO-FORNO

Baseando-se em estudos de dissecação de altos-fornos, conduzidas inicialmente pelos russos e consolidados pelos japoneses, na década de 1970, em que os altos-fornos de diferentes tamanhos, com as mais variadas condições operacionais foram resfriados e dissecados, foi possível definir as cinco principais zonas internas, conforme mostra esquematicamente a Figura 57.

Figura 57 – Zonas internas do Alto-Forno.



Fonte: Apatado de Campos (1984).

A Zona Granular é região onde o minério e o coque descem como carga sólida, mantendo a sua configuração em camadas alternadas, conforme foram carregadas.

Nesta região, ocorre o pré-aquecimento e a secagem da carga sólida e, utilizando a carga térmica e o potencial redutor dos gases provenientes das regiões inferiores, são promovidas reações de redução indiretas. A temperatura desta região varia entre 200° C a 1000° C (GEERDES; TOXOPEUS; VAN DER VLIET, 2009; MOURÃO *et al.*, 2011).

A Zona Coesiva (ou Zona de Amolecimento e Fusão) é a região em que a carga metálica inicia o seu processo de amolecimento e fusão, enquanto que a camada de coque permanece sólida. Este mecanismo aumenta a resistência à passagem do gás, pois o escoamento do fluxo de gases ascendentes é realizado apenas pelas “janelas de coque”. A temperatura desta zona varia entre 1000° C a 1450° C e o seu perfil dependerá basicamente da distribuição da carga e do fluxo gasoso, refletidos pela variação da Relação Minério/Coque no sentido radial do forno (CAMPOS, 1983; MOURÃO *et al.*, 2011).

Neste sentido, conforme aumenta-se a Relação Minério/Coque, menor será a permeabilidade à passagem do gás e vice-versa. Devido à menor circulação de ar, mais tempo será necessário para a carga metálica atingir a temperatura de início de amolecimento, conseqüentemente, a zona coesiva estará localizada em níveis mais baixos. Além disto, quanto maior for o consumo de coque por unidade de tempo, maior será o volume de gás que atravessa a carga, aquecendo-a mais rapidamente. Este aquecimento rápido promove uma elevação da zona coesiva. Portanto, um elevado teor de coque no centro do forno (marcha muito central) ocasiona uma zona em formato “V” invertido, enquanto que um aumento do teor de coque periférico (marcha periférica), acarreta em uma zona com formato “W” (FIGUEIRA; CASTRO; TAVARES, 1989; GEERDES; TOXOPEUS; VAN DER VLIET, 2009).

Abaixo da Zona Coesiva, localiza-se a Zona de Gotejamento, com temperaturas variando entre 1400°C e 1800°C. A medida em que são formados, o gusa e a escória gotejam pelos interstícios do coque sólido, em direção a parte inferior do forno, incorporando os elementos de liga (como o carbono, o manganês, o silício e o fósforo) ao gusa. Esta zona é dividida em duas subzonas: a zona de coque ativo, que compreende a porção do coque que rola em direção às ventaneiras para ser queimado, e o homem morto (*deadman*), a qual é composta pela porção de coque que não rola em direção às ventaneiras e é consumido substancialmente pela incorporação de carbono ao gusa (CAMPOS, 1983; MOURÃO *et al.*, 2011; RIZZO, 2009).

Devido à elevada energia cinética exercida pelo sopro de ar quente das ventaneiras, na Zona de Combustão, forma-se um bolsão de ar (*raceway*) atingindo temperaturas acima de 1900° C. No *raceway*, à medida que as partículas de coque circulam, elas vão sendo queimadas, gerando gás redutor e energia para o processo. O gás resultante sobe através da carga, transferindo calor para a mesma e promovendo as reações de redução (CAMPOS, 1983; MOURÃO *et al.*, 2011).

Na base do alto-forno, próximo ao furo por onde se drena o gusa, localiza-se a Zona do Cadinho. Nesta região, o gusa e a escória são acumulados, separando-se por diferença de densidade, em contato com partículas de coque. No cadinho também ocorrem importantes reações, tal como a dessulfuração do gusa (CAMPOS, 1983; MOURÃO *et al.*, 2011).

## A.2 O FUNCIONAMENTO DO ALTO-FORNO

As matérias-primas ferrosas carregadas pelo topo do alto-forno trazem consigo impurezas como a sílica ( $\text{SiO}_2$ ), o óxido de cálcio ( $\text{CaO}$ ), o óxido de magnésio ( $\text{MgO}$ ) e a alumina ( $\text{Al}_2\text{O}_3$ ), que ao serem fundidas, formarão a escória. Para o controle da proporção adequada dos principais óxidos da escória, utilizam-se os fundentes. Os principais fundentes são: quartzo como fonte de  $\text{SiO}_2$ , calcário como fonte de  $\text{CaO}$ , dunito ou serpentinito como fonte de  $\text{MgO}$  e  $\text{SiO}_2$ , e a dolomita como fonte de  $\text{CaO}$  e  $\text{MgO}$  (CAMPOS, 1983; MOURÃO *et al.*, 2011; RIZZO, 2009).

Na parte inferior do forno, o oxigênio do ar soprado pelas ventaneiras promove a combustão do coque, gerando uma grande quantidade de calor e volume de gases, composto principalmente de nitrogênio ( $\text{N}_2$ ) e monóxido de carbono ( $\text{CO}$ ), que funcionará como o principal redutor dos óxidos de ferro. O ar soprado também contém humidade ( $\text{H}_2\text{O}$ ), que ao reagir com carbono do coque, gera gases redutores, como o  $\text{CO}$  e  $\text{H}_2$  (CAMPOS, 1983; MOURÃO *et al.*, 2011).

Os gases gerados na zona de combustão ( $\text{CO}$ ,  $\text{H}_2$  e  $\text{N}_2$ ), ascendem para o topo do forno através dos vazios existentes entre as partículas, transferindo o calor gerado para a carga que desce em contra-corrente, promovendo a sua secagem, redução, amolecimento e fusão. Desta forma, durante o restante do seu trajeto no forno, o gás ascendente diminui a sua temperatura gradativamente, ao passo que, sua composição química é alterada (ARAUJO, 1997; MOURÃO *et al.*, 2011).

Uma vez que a distribuição de carga afeta diretamente o formato e a posição da zona coesiva e da carga térmica no interior do alto-forno, ela é um dos fatores de maior

relevância para a eficiência das trocas térmicas e das reações de redução. Desta forma, o controle da distribuição da carga é fundamentado a partir dos fluxos gasosos central e periférico.

De acordo com Mourão *et al.* (2011), os gases formados na zona de combustão ascendem rapidamente até o topo (em torno de 6 a 8 segundos). Em contrapartida, levam-se cerca de seis a oito horas para que os materiais sólidos carregados no topo sejam processados e se acumulem na base do alto-forno, na forma de gusa e escória líquidos. Estes produtos líquidos são drenados através dos furos de gusa na região do cadinho, em intervalos de corrida periódicos.

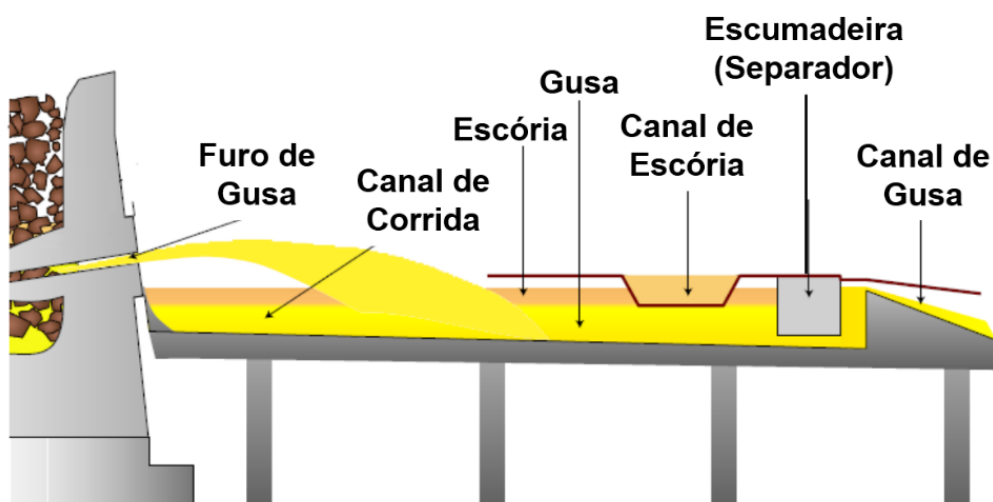
### **A.2.1 O vazamento do gusa**

No cadinho do alto-forno, o ferro-gusa e a escória produzidos se separam por diferença de densidade: a escória possui um peso específico menor ( $2,3 \text{ t/m}^3$ ) do que o gusa ( $7,2 \text{ t/m}^3$ ) e, portanto, flutua sobre ele. No entanto, o volume interno para a acumulação destes materiais é limitado, pois eles ocupam o espaço entre os interstícios do coque presente no cadinho. A acumulação em excesso pode acarretar em impactos operacionais como marcha irregular (baixa produtividade), queima de ventaneiras e esfriamento do cadinho. Portanto, a drenagem regular do cadinho deve ser controlada criteriosamente (MOURÃO *et al.*, 2011; GEERDES; TOXOPEUS; VAN DER VLIET, 2009).

Geerdes, Toxopeus e Van Der Vliet (2009) explicam que, em geral, o alto-forno é vazado de 8 a 14 vezes por dia através do furo de gusa, correspondendo à uma duração média de vazamento entre 90 e 180 minutos. Além disto, o tempo de residência da carga no interior do forno é de aproximadamente 6 horas. Assim, um vazamento de 2 horas, representa um terço do teor do alto forno transformado em ferro e escória. Em grande parte dos fornos de alta produtividade, o vazamento ocorre de maneira alternada e simétrica, assim, no momento em que um furo de gusa é tampado, o outro simetricamente oposto é aberto. Cada vazamento corresponde à uma corrida de gusa e o intervalo de tempo entre as corridas pode ser reduzido para zero (vazamento contínuo) ou, em virtude de desvios operacionais, ocorrer aberturas paralelas.

Normalmente, o primeiro material a ser drenado é o ferro-gusa, logo depois, a escória passa a ser drenada em concomitância com o gusa. Conforme apresenta a Figura 58 (GEERDES; TOXOPEUS; VAN DER VLIET, 2009), o gusa e a escória são separados por diferença de densidade no canal de corrida através de um separador. A escória é então direcionada para um sistema de granulação através do canal de escória, enquanto que o gusa é direcionado pelo canal de gusa para carros torpedos (equipamentos fabricados em aço estrutural e revestido com refratários, cuja função básica é o transporte de ferro-gusa do alto-forno para a aciaria) posicionados no piso inferior da casa de corrida. Um canal basculante é o responsável por direcionar o fluxo de gusa para o carro desejado, permitindo ao operador trocar de carro torpedo durante um vazamento. O final da corrida é caracterizado pelo início da saída de gás do interior do forno pelo furo, apontando que, a superfície de escória naquela região, atingiu o nível do furo de gusa.

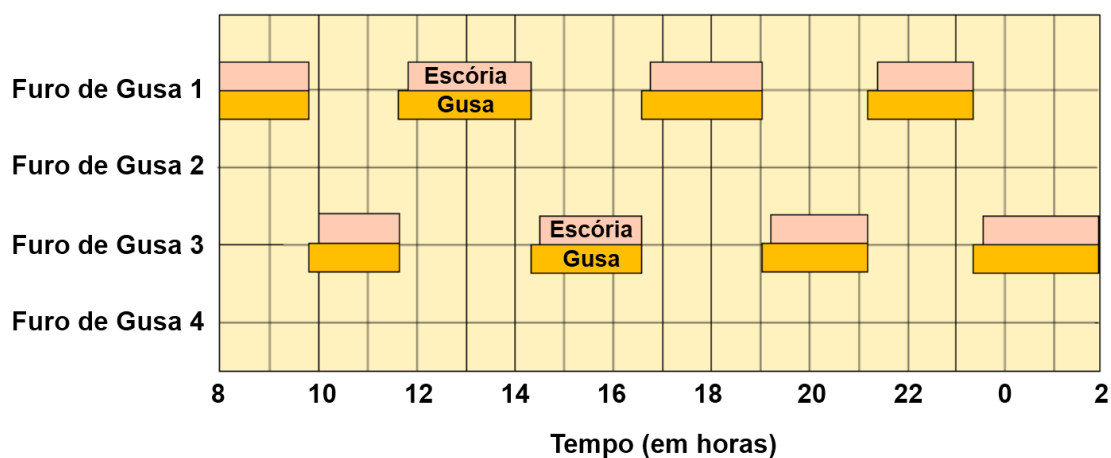
Figura 58 – Vazamento de gusa e escória do cadinho pelo furo de gusa e a separação ao passar pela escumadeira em uma casa de corrida.



Fonte: Adaptado de Geerdes, Toxopeus e Van Der Vliet (2009, p. 125).

Em geral, alto-fornos com cadinhos maiores que 10 metros de diâmetro, possuem 4 furos de gusa. Durante a operação ideal da casa de corrida, o vazamento é feito continuamente, com durações semelhantes para as corridas, alternando-se os furos de gusa em operação. Um padrão típico do fluxo de corridas é ilustrado através da Figura 59 (GEERDES; TOXOPEUS; VAN DER VLIET, 2009).

Figura 59 – Padrão típico de operação de vazamento de um alto-forno com 4 furos de gusa, em que os furos 1 e 3 são operados de forma alternativa e sem atrasos.



Fonte: Adaptado de Geerdes, Toxopeus e Van Der Vliet (2009, p. 129).

Deve-se frisar que, as vazões de ferro-gusa e de escória de um determinado furo de gusa dependem do comprimento e do diâmetro do mesmo, bem como das características físico-químicas dos materiais drenados e da pressão interna do forno. Uma vez que o furo se desgasta continuamente, especialmente pelo ataque da escória, as vazões de gusa e escória não são constantes. Além disso, no início da corrida apenas o gusa poderá ser extraído.

### A.2.2 Controle operacional do Alto-forno

A operação básica do alto-forno consiste do carregamento periódico dos materiais sólidos pelo topo, injeção de ar quente e hidrocarbonetos por meio das ventaneiras, drenagem periódica dos líquidos pela parte inferior e a remoção de gás e poeira pelo topo (MOURÃO *et al.*, 2011).

Visando uma operação estável e com alta produtividade, é fundamental que o controle térmico, a permeabilidade e o esgotamento do cadinho estejam operando em condições adequadas. Desta forma, torna-se necessário o constante monitoramento de parâmetros de controle, muitas vezes obtidos por meio de equipamentos de medição (ARAUJO, 1997; CAMPOS, 1983; MOURÃO *et al.*, 2011).

Embora o processo seja contínuo, ocasionalmente, faz-se necessário a realização de paradas do alto-forno, para a manutenção e reparo de equipamentos. Nestas ocasiões é feita uma preparação de carga do alto-forno com carregamento de maior quantidade de coque. Durante o procedimento de parada ocorrem também a diminuição do

volume de ar (enriquecido com oxigênio) soprado pelas ventaneiras, a redução da injeção de combustível auxiliar e o esgotamento do cadinho. Contudo, deve-se salientar que, dependendo das condições operacionais, podem ocorrer paradas não-programadas ou paradas emergenciais, não havendo tempo hábil para a devida preparação do forno.

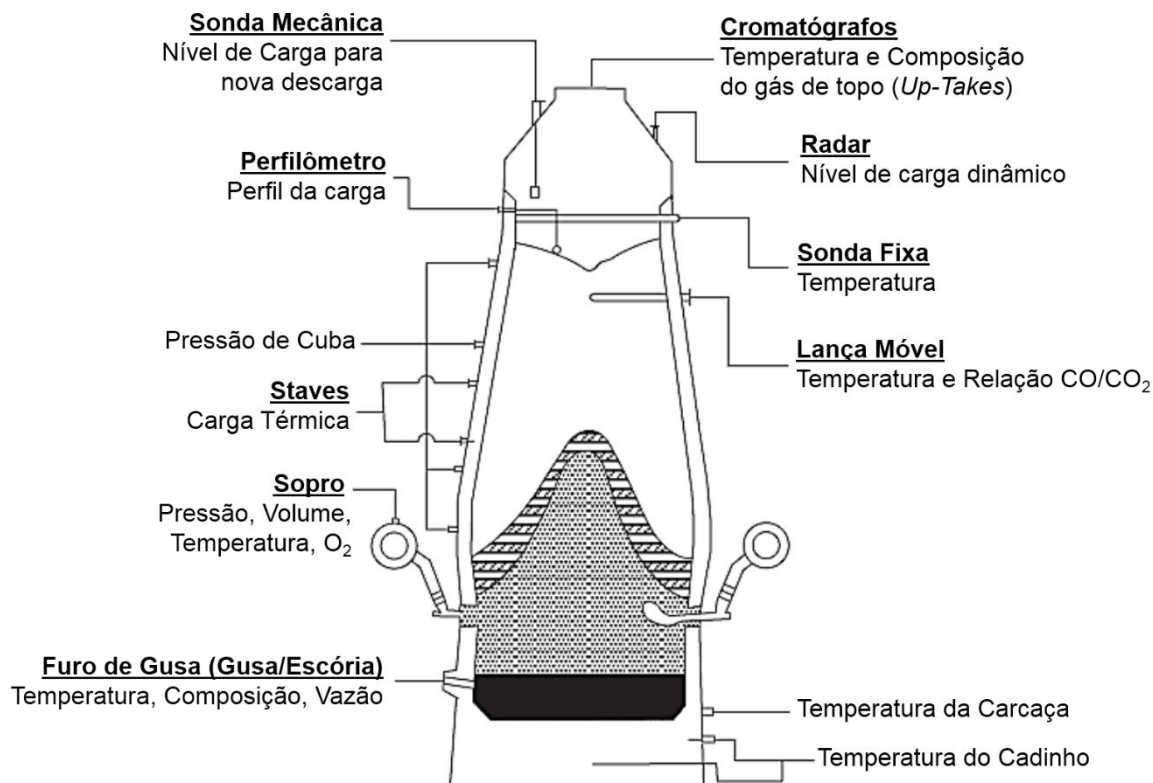
Na prática, variações aleatórias, principalmente as variações nas propriedades das matérias-primas, podem afetar o estado térmico do forno e, portanto, torna-se necessária a constante atuação no controle térmico do processo. Caso contrário, estas variações podem acarretar no comprometimento da qualidade do ferro-gusa, perdas de produção ou até mesmo o resfriamento do cadinho.

Um dos parâmetros mais utilizados para a avaliação do nível térmico do interior do alto-forno é a temperatura do ferro-gusa e as análises de ferro-gusa e escória. Estas análises são realizadas em laboratório, através da espectromia por raio-x, e definem a composição química de cada um dos produtos. O percentual de silício e de enxofre resultantes das análises de ferro-gusa, são excelentes indicadores do nível térmico e atenção deve ser dada ao efeito das matérias-primas nesses teores. A basicidade binária das análises de escória, dada pela razão entre o percentual de CaO e o percentual de SiO<sub>2</sub>, também pode ser levada em consideração na avaliação do aporte térmico devido à sua influência nos níveis de silício no ferro-gusa.

Entretanto, como estes parâmetros servem apenas para a confirmação do estado térmico do alto-forno, eles devem ser tratados como os indicadores dos fenômenos causadores da variação do nível térmico. Portanto, para atuar preventivamente, corrigindo eventuais variações térmicas do interior do alto-forno, os operadores dispõem de algumas alternativas de atuação, baseando-se em variáveis do processo (MOURÃO *et al.*, 2011). A Figura 60 indica a localização de alguns dos equipamentos e indicadores mais relevantes para o controle de processo e funcionamento do forno (WALLACE *et al.*, 1999).



Figura 60 – Representação dos equipamentos e indicadores fundamentais para o controle do processo.



Fonte: Adaptado de Wallace *et al.* (1999, p. 697).

Os Termopares (sensores de temperatura) distribuídos ao longo do forno indicam o nível térmico do forno e fluxo de líquidos no cadinho. Os medidores de pressão monitoram em diferentes elevações do eixo do forno e auxiliam no monitoramento da permeabilidade da carga e a localização da Zona Coesiva. A Lança Móvel é inserida para monitorar a temperatura e a composição do gás no eixo do forno. Estes dados são usados pelos operadores para monitoramento das mudanças na distribuição de carga e eficiência do forno. As Sondas Fixas são empregadas para medir a temperatura do gás, como forma de monitorar padrões de gás através da carga (WALLACE *et al.*, 1999).

A Sonda Mecânica e o Radar monitoram a estabilidade da descida de carga, enquanto que o Perfilômetro indica a distribuição de carga. Nos *Up-Takes* (dutos metálicos que recebem o gás produzido no interior do forno) estão localizados os Cromatógrafos, responsáveis por analisar a composição do gás de topo. A Lança Móvel, a Sonda Fixa e a carga térmica dos *Staves* (placas em contato com a carcaça do alto-forno responsáveis pela refrigeração do mesmo, promovendo uma troca térmica mais

eficiente e aumentando a vida útil do forno) também são responsáveis por evidenciar a marcha do forno, através do fluxo do gás (WALLACE *et al.*, 1999).

Segundo Mourão *et al.* (2011), as principais variáveis utilizadas pelos operadores de alto-forno para o controle do processo são listadas a seguir:

- Taxa de Injeção Auxiliar (Carvão Pulverizado ou Gás Natural): A injeção de carvão pulverizado tem como principal objetivo a redução de custos relativos à produção, por conta da substituição de parte do coque carregado. Esta é uma das principais variáveis de controle térmico devido, essencialmente, ao seu curto tempo de resposta. No entanto, a sua disponibilidade é limitada na faixa inferior pela mínima taxa de injeção e na faixa superior, pela mínima Temperatura de Chama aceitável;
- Temperatura de Sopros: Como forma de minimizar o consumo de coque (Taxa de Coque), procura-se trabalhar com temperaturas mais elevadas. Esta variável também apresenta um curto tempo de resposta;
- Umidade do Ar: Trata-se de uma excelente variável de controle térmico, além de ser utilizada no controle da pressão de base. Na operação com injeções auxiliares, busca-se trabalhar com níveis de umidade mais baixos, para que a temperatura de chama não diminua ainda mais;
- Vazão de Oxigênio (Taxa de Enriquecimento de O<sub>2</sub>): Esta variável está relacionada ao controle de temperatura de chama, que deve ser mantida na faixa de 2000°C a 2400°C, aproximadamente;
- Pressão de Topo: Esta variável está relacionada ao controle da velocidade de gás na cuba, região do alto-forno onde localiza-se a zona granular;
- Taxa de Coque: Para a atuação térmica, a disponibilidade desta variável é limitada pelo longo tempo de resposta. Em geral, gastam-se, no mínimo, cerca de seis horas para que as regiões inferiores do alto forno sejam alcançadas por esta variável;
- Vazão de Ar: Pequenas reduções desta variável são sempre efetivas nos momentos em que o forno estiver termicamente deficiente, mal drenado ou com pressão alta. Estas reduções não apresentam forte impacto negativo no que concerne à produção do forno.

Uma vez que, a Temperatura de Chama é influenciada pelas condições qualitativas da zona de combustão e, por esta razão, ela também é conhecida como Temperatura

do *Raceway*. Portanto, ela irá aumentar se a temperatura de sopro e o percentual de oxigênio aumentarem, e diminuirá se a umidade e a injeção de combustível auxiliar aumentarem. Em relação à produtividade do alto-forno (marcha do forno), ela é determinada pela quantidade de oxigênio soprado através das ventaneiras, a partir de um determinado quantitativo de coque e carvão usado por tonelada de ferro-gusa (*Fuel Rate*). Desta forma, quanto maior for a quantidade de oxigênio soprado pelas ventaneiras, maior será a quantidade de CO resultante da combustão do coque e mais alta será a taxa de produção (GEERDES; TOXOPEUS; VAN DER VLIET, 2009).

Neste sentido, os operadores de alto-forno são capazes de determinar o estado térmico do alto-forno a partir de suas variáveis controle, medidas muitas vezes por instrumentação padrão ou calculadas a partir de outros índices do sistema. Através destas informações, os operadores corrigem eventuais anormalidades do processo, garantindo assim a estabilidade e a qualidade do produto final. No entanto, em decorrência dos atrasos de tempo entre as ações de controle, a correlação existente entre as próprias variáveis do forno, bem como a complexidade do processo, o controle do aporte térmico não é uma tarefa trivial.

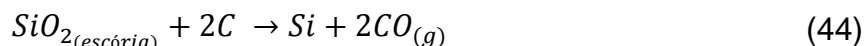
### A.3 O TEOR DE SILÍCIO NO FERRO-GUSA

De uma maneira geral, dentre os elementos de liga que compõem o ferro-gusa, o enxofre e o silício são os mais importantes, conseqüentemente, são os elementos nos quais os operadores procuram exercer maior controle. No caso do enxofre, o seu controle torna-se necessário devido aos seus efeitos nocivos às propriedades mecânicas do aço. Com relação ao silício, sua importância é devido a sua capacidade em expressar não somente o estado térmico do forno, como também a qualidade do produto final e os custos associados à produção (FIGUEIRA; CASTRO; TAVARES, 1989; MOURÃO *et al.*, 2011).

Neste contexto, as variações no nível térmico do interior do alto-forno são refletidas em flutuações do teor de silício do ferro-gusa. O teor do silício depende não só da quantidade como também dos tipos de matérias-primas empregadas (minério e coque) e na forma como foi conduzido o processo de fabricação do gusa (MOURÃO *et al.*, 2011).

### A.3.1 Mecanismos de transferência do silício

A principal fonte de silício no alto-forno provém da sílica ( $\text{SiO}_2$ ) presente nos minérios, das cinzas do coque e do carvão mineral pulverizado pelas ventaneiras. A transferência do silício para o gusa pode ocorrer através de dois mecanismos: diretamente da escória para o gusa, de acordo com a Equação (44), através de reações metal-escória,



e indiretamente, de maneira que o silício é transferido da escória (ou da cinza do coque) para a fase gasosa ( $\text{SiO}$ ), que então reage com o carbono do ferro-gusa, incorporando o silício, conforme apresentam as Equações (45) e (46) (GEERDES; TOXOPEUS; VAN DER VLIET, 2009; MOURÃO *et al.*, 2011).



Como a Equação (44) é altamente dependente da temperatura (endotérmica), quanto maior o nível térmico do forno, maior a temperatura do gusa e, conseqüentemente, maior a probabilidade de incorporação de silício. Figueira, Castro e Tavares (1989) e Mourão *et al.* (2011) explicam que, estudos cinéticos da reação metal-escória de vários autores, concluíram que a velocidade desta reação na zona de gotejamento é bastante lenta, conseqüentemente, a transferência de silício via  $\text{SiO}$ , das Equações (45) e (46), é o principal mecanismo de transferência.

A reação descrita pela Equação (45) é fortemente influenciada pelo aumento da Temperatura de Chama e diminuição da Pressão de Sopro das ventaneiras. Já a reação descrita pela Equação (46) sofre influência da elevação da Zona Coesiva, refletida pelo aumento da Zona de Gotejamento. Esta elevação faz com que se prolongue o tempo disponível para que esta reação aconteça.

Neste contexto, pode-se afirmar que dois dos fatores mais afetam a incorporação de silício no gusa é a temperatura do gusa e a posição e geométrica da Zona Coesiva. De uma maneira geral, considerando que, ambos os fatores são diretamente proporcionais à quantidade de combustível inserido ao processo, quanto maior for o teor de silício, maior será o consumo de combustíveis (*Fuel Rate*), principalmente do coque (Taxa de Coque). Assim, baixos teores de silício refletem uma operação eficiente, indicando uma possível economia de combustível.

Geerdes, Toxopeus e Van Der Vliet (2009), Mourão *et al.* (2011) e Figueira, Castro e Tavares (1989) destacam que existem outras condições de oxidação silício do gusa quando em contato com a escória. Após passar pelo nível das ventaneiras, o teor de silício do gusa é reduzido devido a sua oxidação pelo MnO e do FeO da escória no cadinho (Equações (47) e (48), respectivamente), sendo a última reação a mais intensa.



Em virtude da presença de carbono misturado à escória, algumas reações escória-coque podem ocorrer, conforme descrevem as Equações (49), (50) e (51):



No entanto, um aumento da Pressão de Sopro pode descolar o equilíbrio das reações de redução do MnO e do FeO, de maneira que as reações dadas pelas Equações (52) e (53), consumam menores quantidades destes óxidos. Assim, a quantidade de agentes oxidantes do silício aumentam, resultando em teores de silício mais baixos.



Outro aspecto relevante destacado pelos autores é que quanto mais básica for a escória, isto é, quanto mais baixo for a quantidade de SiO<sub>2</sub> nela, menor será o teor de silício no gusa.

Em virtude de todas estas reações, durante a produção de gusa com alto teor de silício, é comum observar a presença de SiO nos gases, através de um pó branco na poeira do topo. Além disto, Geerdes, Toxopeus e Van Der Vliet (2009) destacam que, os valores de silício no gusa após uma parada ou partida de alto-forno são extremamente elevados, especialmente se, durante estes procedimentos o forno estiver operando com um baixo Volume de Sopro.

Portanto, observa-se que diversos fatores durante o processo podem afetar no percentual de silício incorporado ao gusa. Contudo, muitos destes fatores são correlacionados entre si, de forma que o efeito simultâneo de todos eles é de difícil previsão (FIGUEIRA; CASTRO; TAVARES, 1989).

### A.3.2 Efeitos do silício no processo produtivo

Tradicionalmente, o teor de silício no ferro-gusa é utilizado pelos operadores como principal indicador do estado térmico do alto-forno. Um baixo teor de silício indica um arrefecimento do forno, podendo comprometer sua operação, enquanto que altos valores indicam um desperdício de coque, evidenciado através da excessiva geração de calor. Dado que o custo do coque é dominante no processo, a economia deste combustível pode refletir em benefícios econômicos bastante evidentes na fabricação do ferro-gusa (MOURÃO *et al.*, 2011; RIZZO, 2009; GEERDES; TOXOPEUS; VAN DER VLIET, 2009; SAXEN; PETTERSSON, 2007; SAXEN; GAO, C.; GAO, Z., 2013). Ao sair do alto-forno, o gusa líquido é levado à aciaria, onde é refinado e transformado em aço, através dos convertedores. O aporte térmico do gusa é o principal responsável pelo fornecimento de calor ao processo e, por isso, este material deve apresentar características físico-químicas adequadas e com a menor variação possível, caso contrário, a produtividade é comprometida em virtude da constante necessidade de ajustes ao processo. Logo, um alto teor de silício pode prejudicar o processo de conversão do ferro em aço e, conseqüentemente, gerar gastos ainda maiores.

Na aciaria, o teor de silício no ferro-gusa é determinante para a operação regular dos convertedores. Teores muito altos de silício no gusa aumentam o volume de escória gerada pelo processo, aumentando-se o risco de ocorrência de projeções para fora do convertedor. Ao entrar em contato com o solo, a escória gera uma fumaça de cor avermelhada, ocasionando emissão ambiental.

Além disto, muitas vezes é necessário que o sopro de oxigênio seja interrompido para que ocorra o vazamento de parte da escória, caracterizando o efeito da dupla escória. A dupla escória consiste em um duplo derramamento de escória no processo dos convertedores. Em virtude da necessidade da diminuição do aporte térmico por conta dos elevados teores de silício, é necessária a inserção de material refrigerante. Em decorrência desta adição, ocorre uma maior produção de escória, devido à maior oxidação do silício.

A dupla escória diminui a produtividade, uma vez que, o processo precisa ser interrompido. Além disto, aumenta-se o gasto na produção do aço, uma vez que, torna-se necessário uma maior quantidade de matéria-prima, como fundentes e materiais refrigerantes. Todos estes fatores afetam o rendimento do sistema de conversores e a sua produtividade, em decorrência da constante necessidade de

ajustes ao processo (CAMPOS, 1983; GUSTAVSSON, 2004; MOURÃO *et al.*, 2011; RIZZO, 2009).

Em contrapartida, a concentração do silício no gusa líquido representa boa parte da energia necessária para o sistema de convertedores, uma vez que a oxidação do silício promove uma reação exotérmica. Por esta razão, baixos teores de silício alteram as condições de oxidação, prejudicando a formação de escória no início do processo, aumentando a eminência de projeções de metal para fora do convertedor. Nestes casos, faz-se necessária a inserção de Carbetos de Silício (SiC) para que promova reações de redução da sílica (SiO<sub>2</sub>), no entanto, isso faz com que se aumente os custos associados ao processo (CAMPOS, 1983; GUSTAVSSON, 2004).

## ANEXO A: MODELOS ARIMA E ARIMAX

A classe de modelos Autorregressivos, Integrados e de Médias Móveis (ARIMA) foi sistematizada por Box e Jenkins (1976) e é considerada uma das classes de modelos de previsão para séries temporais mais populares. Conforme está representado na Figura 61, estes modelos são baseados no modelo de filtro linear, o qual supõe que a série temporal  $Z_t$  é gerada através de um filtro com função de transferência  $\psi(B)$ , cuja a entrada trata-se de um ruído branco  $a_t$  no tempo  $t$  (BARROS, 2004; MORETTIN; TOLOI, 2006).

Figura 61 – Filtro linear, com entrada  $a_t$ , saída  $Z_t$  e função de transferência  $\psi(B)$ .



Fonte: O próprio autor.

Na modelagem proposta por Box e Jenkins (1994), o filtro linear é descrito pela Equação (54), como a razão de dois polinômios,

$$\psi(B) = \frac{\theta_q(B)}{\phi_p(B)} \quad (54)$$

sendo  $B$  o operador de atraso, cujo efeito sobre a variável discreta é  $B^k Z_t = Z_{t-k}$ , e  $\theta_q(B)$  e  $\phi_p(B)$  são polinômios de graus  $q$  e  $p$ , que representam, respectivamente, os operadores de Média Móvel (MA( $q$ )) e Autorregressivos (AR( $p$ )). Seguindo este raciocínio, os primeiros modelos Box e Jenkins são definidos pela Equação (55):

$$\phi_p(B)Z_t = \theta_q(B)a_t \quad (55)$$

Esta classe de modelos é denominada modelo Autorregressivo de Média Móvel (ARMA). O modelo ARMA( $p, q$ ) também pode ser representado através da Equação (56):

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \theta_1 a_{t-1} + \theta_q a_{t-q} + a_t \quad (56)$$

no qual  $Z_{t-p}$  corresponde ao valor da série temporal  $p$  instantes atrás,  $\phi_p$  é o parâmetro autorregressivo relativo à amostra  $Z_{t-p}$ , e  $\theta_q$ , o parâmetro de média móvel referente ao ruído  $a_{t-q}$ .

No entanto, os modelos ARMA são apenas um caso particular de um modelo ARIMA. Uma vez que os modelos ARMA são capazes de representar apenas modelos



estacionários, é necessário tomar um número finito de diferenças de  $Z_t$ , para lidar com modelos não-estacionários. Desta forma, considera-se  $W_t$  uma série diferenciada e estacionária, conforme ilustra a Equação (57):

$$W_t = \Delta^d Z_t \quad (57)$$

sendo  $\Delta^d$ , a diferença de ordem  $d$ , é possível representar  $W_t$  por um modelo ARMA e ao integrá-lo, obter  $Z_t$ . Esta é a definição de um modelo ARIMA( $p, d, q$ ) e pode ser descrito pela Equação (58):

$$\phi_p(B)\Delta^d Z_t = \theta_q(B)a_t \quad (58)$$

A vantagem desta classe de modelos ARIMA( $p, d, q$ ) é que eles podem ser aplicados a séries estacionárias ou não. A condição necessária para a sua aplicação em séries não-estacionárias é a possibilidade da estacionariedade ser obtida, por exemplo, através de diferenças da série original (MORETTIN; TOLOI, 2006).

Em resumo, as propriedades da família ARIMA( $p, d, q$ ) dependem dos inteiros não-negativos  $p$ ,  $d$  e  $q$  responsáveis por determinar a ordem do modelo dentro da classe. Seguindo esta notação,  $p$  é o número de termos autorregressivos,  $d$  é o número de diferenças e  $q$  é o número de termos de médias móveis. Deve-se ressaltar que, caso a série seja estacionária, não haverá termo  $d$  e, portanto, o ARIMA se reduzirá a um modelo ARMA (BARROS, 2004).

Os modelos ARMA que incluem outras séries temporais como variáveis de entrada, são chamados de modelos Autorregressivos de Médias Móveis com Entradas Exógenas (ARMAX). Estes modelos são similares aos modelos ARMA, porém, adicionalmente, possuem entradas exógenas. Desta forma, eles podem ser simbolizados por ARMAX( $p, q, r_x$ ), em que  $r_x$  corresponde à ordem da equação que representa a variável exógena. Esta classe de modelos tem a capacidade de identificar os padrões subjacentes nos dados da série temporal e quantificar o impacto de influências externas, representadas pelas variáveis observáveis. Este modelo é caracterizado pela Equação (59),

$$\begin{aligned} Z_t = & \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_{r_x} X_{t-r_x} + \dots \\ & \dots + \theta_1 a_{t-1} + \theta_q a_{t-q} + a_t \end{aligned} \quad (59)$$

na qual  $X_{t-r_x}$  corresponde ao valor da entrada exógena  $r_x$  instantes atrás e  $\beta_{r_x}$  o seu respectivo coeficiente.

No caso da série temporal não ser estacionária, é necessário realizar a mesma quantidade de diferenciações tanto na saída quanto na entrada. Neste caso, empregam-se os modelos Autorregressivos, Integrados de Médias Móveis com Entradas Exógenas (ARIMAX), que assumem a forma da Equação (60):

$$\phi_p(B)\Delta^d Z_t = \beta_{r_x}(B)\Delta^d X_t + \theta_q(B)a_t \quad (60)$$

Deve-se ressaltar que, o modelo ARIMAX pode ser generalizado para os casos em que mais de uma entrada exógena é necessária para a modelagem do sistema (BOX; JENKINS; REINSEL, 1976; HAMILTON, 1994; PANKRATZ, 1991).

## ANEXO B: RESULTADOS OBTIDOS APÓS CADA IMPLEMENTAÇÃO DO ALGORITMO DE PODA APLICADO À REDE NIO

Tabela 18 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NIO (Continua).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
1	Relação CO/CO <sub>2</sub>	1	4	0,00502	18,61	54,91
	Eficiência do gás H <sub>2</sub>	3				
	Taxa de Coque	5				
	<i>PCI Rate</i>	3				
	Relação Minério/Coque	2				
	IFGCN	2				
	Volume de Sopro	1				
	Taxa de O <sub>2</sub>	2				
	Pressão de Sopro	1				
	Teor de CO <sub>2</sub> de Topo	2				
	Temperatura do ferro-gusa	2				
2	Taxa de Coque	6	5	0,00491	18,42	55,47
	<i>PCI Rate</i>	2				
	Índice Térmico H0	3				
	Umidade de Sopro	2				
	Temperatura do ferro-gusa	2				
3	Taxa de Coque	6	5	0,00568	20,08	51,40
	<i>PCI Rate</i>	2				
	Taxa de O <sub>2</sub>	2				
	Teor de CO <sub>2</sub> de Topo	2				
	Eficiência do gás CO	2				
	Temperatura do ferro-gusa	2				

Tabela 18 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NIO (Continuação).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
4	Relação CO/CO <sub>2</sub>	1	5	0,00492	18,19	53,93
	Taxa de Coque	6				
	Relação Minério/Coque	2				
	IFGCS	2				
	Taxa de O <sub>2</sub>	1				
	Teor de CO <sub>2</sub> de Topo	2				
5	Temperatura de Chama	1, 3	8	0,00482	18,49	51,96
	Relação CO/CO <sub>2</sub>	1, 3				
	Eficiência do gás H <sub>2</sub>	2				
	Velocidade Produção de Gusa	3				
	Taxa de Coque	1, 2, 3, 5, 6, 7				
	PCI Rate	1, 2				
	Relação Minério/Coque	1, 2, 7				
	Índice Térmico H0	3				
	IFGCN	1, 2				
	IFGPN	1, 2				
	IFGCS	1, 2				
	IFGPS	1				
	Umidade de Sopro	1, 3				
	Taxa de O <sub>2</sub>	2, 3				
	Temperatura de Sopro	2				
	Pressão de Sopro	2				
	Temperatura do gás de Topo	1, 2				
	Teor de H <sub>2</sub> de Topo	3				
	Teor de N <sub>2</sub> de Topo	2, 3				

Tabela 18 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NIO (Continuação).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
5	Teor de CO de Topo	2, 3	8	0,00482	18,49	51,96
	Teor de CO <sub>2</sub> de Topo	2				
	Eficiência do gás CO	1, 3				
	Perda Térmica <i>Stave</i>	1, 2				
	Temperatura do ferro-gusa	2, 3				
	Força-Eletromotriz	1, 2, 3				
6	Temperatura de Chama	1	7	0,00553	19,79	53,79
	Eficiência do gás H <sub>2</sub>	3				
	Taxa de Coque	5				
	<i>PCI Rate</i>	1, 3				
	IFGPS	3				
	Volume de Sopro	1, 2				
	Pressão de Sopro	2				
	Temperatura do ferro-gusa	1				
7	Temperatura de Chama	2	6	0,00553	19,99	51,12
	Relação CO/CO <sub>2</sub>	2				
	Velocidade Produção de Gusa	1				
	Taxa de Coque	5				
	<i>PCI Rate</i>	3				
	Taxa de O <sub>2</sub>	2				
	Temperatura do ferro-gusa	2				

Tabela 18 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NIO (Continuação).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
	Temperatura de Chama	1, 3				
	Relação CO/CO <sub>2</sub>	3				
	Eficiência do gás H <sub>2</sub>	1, 3				
	Velocidade Produção de Gusa	1				
	Taxa de Coque	3, 5, 6, 7				
	<i>PCI Rate</i>	3				
	Redução Direta	2, 3				
	Relação Minério/Coque	2				
	IFGCN	2				
	IFGPN	2, 3				
	IFGCS	2, 3				
	IFGPS	1, 2, 3				
8	Volume de Sopro	1	9	0,00422	17,40	58,42
	Umidade de Sopro	2, 3				
	Taxa de O <sub>2</sub>	2, 3				
	Temperatura de Sopro	1				
	Pressão de Sopro	1, 3				
	Pressão de Topo	2				
	Temperatura do gás de topo	1, 3				
	Teor de N <sub>2</sub> de Topo	1, 2				
	Teor de CO de Topo	2, 3				
	Teor de CO <sub>2</sub> de Topo	2				
	Eficiência do gás CO	1, 3				
	Perda Térmica <i>Stave</i>	1, 3				

Tabela 18 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NIO (Conclusão).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
8	Temperatura do ferro-gusa	1, 2, 3	9	0,00422	17,40	58,42
	Força-Eletromotriz	1, 2				
9	Taxa de Coque	5	3	0,00541	19,10	56,60
	Taxa de O <sub>2</sub>	1				
	Temperatura do ferro-gusa	2				
10	Temperatura de Chama	3	6	0,00516	19,43	54,36
	Relação CO/CO <sub>2</sub>	2				
	Taxa de Coque	6				
	PCI Rate	1, 2				
	Redução Direta	2, 3				
	Relação Minério/Coque	4				
	IFGCN	1				
	IFGPN	2				
	Volume de Sopro	1, 2				
	Taxa de O <sub>2</sub>	2, 3				
	Temperatura de Sopro	1, 2				
	Pressão de Sopro	3				
	Temperatura de gás de topo	3				
	Teor de H <sub>2</sub> de Topo	2				
	Teor de CO de Topo	2, 3				
	Teor de CO <sub>2</sub> de Topo	2				
	Temperatura do ferro-gusa	2, 3				
	Força-Eletromotriz	1, 3				

Fonte: O próprio autor.

## ANEXO C: RESULTADOS OBTIDOS APÓS CADA IMPLEMENTAÇÃO DO ALGORITMO DE PODA APLICADO À REDE NARX

Tabela 19 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NARX (Continua).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
1	Saída Autorregressiva	1, 2, 3, 4, 5	3	0,0089	7,21	92,97
	Velocidade de Produção de Gusa	2				
	Relação Minério/Coque	3				
	Índice Térmico H0	2				
	IFGP N	2				
	Taxa de O <sub>2</sub>	1				
	Temperatura de Sopro	2				
2	Força Eletromotriz	2	10	0,0011	7,32	91,292
	Saída Autorregressiva	1, 2, 3, 4, 5				
	Temperatura de Chama	1, 2, 3				
	Relação CO/CO <sub>2</sub>	1, 2, 3				
	Eficiência do gás H <sub>2</sub>	1, 2, 3				
	Velocidade de Produção de Gusa	1, 2, 3				
	Taxa de Coque	1, 2, 3, 4, 5, 6, 7				
	PCI Rate	1, 2, 3				
	Redução Direta	1, 2, 3				
	Relação Minério/Coque	1, 2, 3, 4, 5, 6, 7				
	Índice Térmico H0	1, 2, 3				
	IFGCN	1, 2, 3				
	IFGPN	1, 2, 3				
	IFGCS	1, 2, 3				
	IFGPS	1, 2, 3				
	Volume de Sopro	1, 2, 3				



Tabela 19 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NARX (Continuação).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
2	Umidade de Sopro	1, 2, 3	10	0,0011	7,32	91,292
	Taxa de O <sub>2</sub>	1, 2, 3				
	Temperatura de Sopro	1, 2, 3				
	Pressão de Topo	1, 2, 3				
	Temperatura do gás de Topo	1, 2, 3				
	Teor de H <sub>2</sub> de Topo	1, 2, 3				
	Teor de N <sub>2</sub> de Topo	1, 2, 3				
	Teor de CO de Topo	1, 2, 3				
	Teor de CO <sub>2</sub> de Topo	1, 2, 3				
	Eficiência do gás CO	1, 2, 3				
	Perda Térmica Stave	1, 2, 3				
	Temperatura do ferro-gusa	1, 2, 3				
	Força-Eletromotriz	1, 2, 3				
3	Saída Autorregressiva	1, 2, 3, 4, 5	5	0,0008	7,35	92,696
	Temperatura de Chama	1, 2, 3				
	Relação CO/CO <sub>2</sub>	1, 2, 3				
	Eficiência do gás H <sub>2</sub>	1, 2, 3				
	Velocidade de Produção de Gusa	1, 2, 3				
	Taxa de Coque	1, 2, 3, 4, 5, 6, 7				
	PCI Rate	1, 2, 3				
	Redução Direta	1, 2, 3				
	Relação Minério/Coque	1, 2, 3, 4, 5, 6, 7				
	Índice Térmico H0	1, 2, 3				

Tabela 19 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NARX (Continuação).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
3	IFGCN	1, 2, 3	5	0,0008	7,35	92,696
	IFGPN	1, 2, 3				
	IFGCS	1, 2, 3				
	IFGPS	1, 2, 3				
	Volume de Sopro	1, 2, 3				
	Umidade de Sopro	1, 2, 3				
	Taxa de O <sub>2</sub>	1, 2, 3				
	Temperatura de Sopro	1, 2, 3				
	Pressão de Topo	1, 2, 3				
	Temperatura do gás de Topo	1, 2, 3				
	Teor de H <sub>2</sub> de Topo	1, 2, 3				
	Teor de N <sub>2</sub> de Topo	1, 2, 3				
	Teor de CO de Topo	1, 2, 3				
	Teor de CO <sub>2</sub> de Topo	1, 2, 3				
	Eficiência do gás CO	1, 2, 3				
	Perda Térmica <i>Stave</i>	1, 2, 3				
	Temperatura do ferro-gusa	1, 2, 3				
	Força-Eletromotriz	1, 2, 3				
4	Saída Autorregressiva	1, 2, 3, 4, 5	8	0,0010	7,79	90,59
	Temperatura de Chama	1, 3				
	Eficiência do gás H <sub>2</sub>	1, 2				
	Velocidade de Produção de Gusa	3				
	Taxa de Coque	1, 3, 5				
	<i>PCI Rate</i>	3				

Tabela 19 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NARX (Continuação).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
4	Relação Minério/Coque	1, 3	8	0,0010	7,79	90,59
	Índice Térmico H0	1, 2, 3				
	IFGCN	2, 3				
	IFGPN	2, 3				
	IFGCS	1				
	IFGPS	1, 3				
	Volume de Sopro	2, 3				
	Umidade de Sopro	1, 2, 3				
	Taxa de O <sub>2</sub>	2, 3				
	Temp. de Sopro	1, 3				
	Pressão de Topo	1, 3				
	Temperatura do gás de Topo	1				
	Teor de H <sub>2</sub> de Topo	2				
	Teor de CO <sub>2</sub> de Topo	1				
	Eficiência do gás CO	1				
	Perda Térmica <i>Stave</i>	1, 2, 3				
	Temperatura do ferro-gusa	2				
	Força-Eletromotriz	2, 3				
5	Saída Autorregressiva	1, 2, 3, 4, 5	3	0,0008	6,32	93,11
	Temperatura do gás de Topo	1				
6	Saída Autorregressiva	1, 2, 3, 4, 5	4	0,0010	7,65	89,32
7	Saída Autorregressiva	1, 2, 3, 4, 5	6	0,0011	8,00	89,18
	IFGC S	1, 3				
	Volume de Sopro	1				

Tabela 19 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NARX (Continuação).

Poda	Entrada	Ordem do Regressor	Número de Neurônios Ocultos	MSE	MAPE (%)	PEA (%)
7	Temperatura de Sopros	3	6	0,0011	8,00	89,18
	Temperatura do gás de Topo	1				
	Temperatura do ferro-gusa	2				
8	Saída Autorregressiva	1, 2, 3, 4	7	0,0007	6,89	92,13
	Eficiência do gás H <sub>2</sub>	2				
	Taxa de Coque	7				
	PCI Rate	3				
	Índice Térmico H0	3				
	IFGPN	3				
	IFGCS	1, 3				
	IFGPS	2				
	Umidade de Sopros	1				
	Temperatura de Sopros	2				
	Teor de CO de Topo	2				
	Eficiência do gás CO	1				
	Perda Térmica Stave	2				
9	Saída Autorregressiva	1, 2, 3, 4, 5	6	0,0011	6,41	93,82
	Relação CO/CO <sub>2</sub>	2				
	Eficiência do gás H <sub>2</sub>	3				
	Velocidade de Produção de Gusa	3				
	Taxa de Coque	1, 2, 3				
	Redução Direta	3				
	Relação Minério/Coque	1				
	IFGPN	1, 2				

Tabela 19 – Resultados obtidos após cada implementação do Algoritmo de Poda aplicado à Rede NARX (Conclusão).

<b>Poda</b>	<b>Entrada</b>	<b>Ordem do Regressor</b>	<b>Número de Neurônios Ocultos</b>	<b>MSE</b>	<b>MAPE (%)</b>	<b>PEA (%)</b>
9	IFGCS	2	6	0,0011	6,41	93,82
	Volume de Sopro	1				
	Umidade de Sopro	1				
	Taxa de O <sub>2</sub>	3				
	Temperatura do gás de Topo	2				
	Teor de CO <sub>2</sub> de Topo	2				
	Eficiência do gás CO	2				
	Temperatura do ferro-gusa	1, 2				
	Força-Eletromotriz	1				
10	Saída Autorregressiva	1, 2, 3, 4, 5	3	0,0008	7,36	92,696
	Taxa de Coque	3				

Fonte: O próprio autor.